Talking science: the research evidence on the use of small-group discussions in science teaching

Bennett, Judith; Hogarth, Sylvia; Lubben, Fred; Campbell, Bob; Robinson, Alison

Talking science: the research evidence on the use of small-group discussions in science teaching

Journal:	*International Journal of Science Education*
Manuscript ID:	TSED-2008-0099.R2
Manuscript Type:	Research Paper
Keywords:	learning activities, literature review
Keywords (user):	small group discussion

Talking science: the research evidence on the use of small-group discussions in science teaching

Abstract

This paper reports the findings of two systematic reviews of the use and effects of small-group discussions in high school science teaching. 94 studies were included in an overview (systematic map) of work in the area, and 24 studies formed the basis of the in-depth reviews.

The reviews indicate that there is considerable diversity in the topics used to promote small-group discussions. They also demonstrate that students often struggle to formulate and express coherent arguments, and demonstrate a low level of engagement with tasks. The reviews suggest that groups function more purposefully, and understanding improves most, when specifically constituted such that differing views are represented, when some form of training is provided for students on effective group work, and when help in structuring discussions is provided in the form of 'cues'. Single sex groups function more purposefully than mixed sex groups, though improvements in understanding are independent of gender composition of groups. Finally, the reviews demonstrate very clearly that, for small-group discussions to be effective, teachers and students need to be given explicit teaching in the skills associated with the development of arguments and the characteristics associated with effective group discussions.

In addition to the substantive findings, the paper also reports on key features of the methods employed to gather and analyse data. Of particular note are the two contrasting approaches to data analysis, one adopting a grounded theory approach and the other drawing on established methods of discourse analysis.

Introduction

The use of small-group discussions in teaching has been advocated for a number of years in science teaching, both to motivate students and to enhance their learning. Such tasks are now appearing with greater frequency in a range of high school science teaching resources. Despite this increasing popularity, comparatively little is known about the detail of their use and effects. This paper examines the research evidence on the use of small-group discussions through presenting the findings of two systematic reviews, with a view to making recommendations for policy and practice on the use of small-group discussions in science teaching.

As originally conceived, the reviews had two principal aims. The first of these was to identify the ways in which small-group discussions are currently used in science lessons. The second was to look at the effects of small group discussions on students' understanding of science and on students' attitudes to science. In practice, the review work established that there is a dearth of studies reporting in any detailed and systematic way on the effects of small-group discussions on students' attitudes to science, so this paper reports the review findings on the use of small-group discussions in science teaching and their effects on students' understanding of science ideas.

Several factors have contributed to the current high levels of interest in small-group discussion work in science. The use of small-group discussion in teaching has its origins in learner-centred teaching approaches, and is one of a range of 'active learning' strategies aimed at stimulating students' interest in what they are studying by providing them with a significant degree of autonomy over the learning activity (e.g. Bentley and Watts, 1992; Kyriacou, 1998). A number of people have advocated the use of discussions in science lessons. Lemke (1990) argues that "learning science means learning to talk science" (p1), and that this means moving away from science lessons dominated by teacher talk in which the teacher asks a question, then invites and evaluates a student response. Lemke refers to this as 'triadic dialogue', which is similar to the initiation-response-feedback (IRF) sequence characterised earlier by Sinclair and Coulthard (1975). Lemke and others (e.g. Sutton, 1992, and Wellington and Osborne, 2001) criticise the approach for leading to talk that focuses on what the teacher wants to hear, rather than promoting genuine communication, and all argue for increased use of discussion work in science lessons.

Support for the use of small-group discussions in science teaching has also emerged in the recommendations from other areas of work in science education. For example, research on alternative conceptions has explored in depth the ideas and understanding students bring with them to science lessons and the ways in which some of their ideas may hinder the development of accepted scientific ideas (e.g. Driver *et al.*, 1985). Small-group discussions have been suggested as a means of helping students explore their ideas and move from understandings that may often be naïve to towards more valid scientific ideas and explanations. Further impetus for the inclusion of small-group discussions in science lessons has come from the development of ideas about *social constructivism* (Driver *et al.*, 1994), which draws on the work of Vygotsky and the importance to learning of social dynamics of interaction (Scott, 1998).

The publication of *Beyond 2000* (Millar and Osborne, 1998) in the UK stimulated discussion and debate over the nature of the school science curriculum and, in particular, the ways in which it might foster the development of *scientific literacy*. This term embraces the knowledge, understanding and skills young people need to develop in order to think and act appropriately on scientific matters which may affect their lives and the lives of other members of the local, national and global communities of which they are a part. Many of the materials developed include small-group discussions in the repertoire of activities employed in science lessons in order to encourage students to participate in informed discussion and debate of scientific issues (e.g. Millar, 2006). Linked to the development of scientific literacy is that of *ideas about evidence* (Millar and Osborne, 1998), which involves encouraging students to evaluate, interpret and analyse evidence from primary and secondary sources in science, including stories about how important science ideas were first developed, and then established and finally accepted. This has led to considerations of the role of *argument* in school science, in the sense of putting forward claims and supporting them with sound and persuasive evidence (e.g. Newton *et al.*, 1999; Osborne *et al.*, 2001). Small-group discussions are viewed as having a key role to play here, since the practice of using evidence in argumentation requires interaction with peers.

In the UK, two other developments have also served to raise the profile of small-group discussions. Firstly, the publication of *Inside the Black Box* (Black and Wiliam, 1998) has resulted in considerable attention being paid to formative assessment, or assessment for learning. Small-group discussions are one approach that has been advocated for increasing the use and effectiveness of formative assessment in science teaching (see, for example, Daws and Singh, 1999). Secondly, there is a more general drive to improve students' *literacy skills*. In England and Wales, this has been formalised into the National Literacy Strategy (DfEE, 1998), where small-group discussions have been advocated as a means for developing students' oral communication skills in science.

Whilst the preceding discussion point to a general recognition of the potential value of small-group discussion in science teaching, their use challenges the established pedagogy of science teaching and place new demands on teachers (see, for example, Bentley and Watts, 1992). There is a growing body of evidence, both anecdotal and more systematic, that many teachers lack skills and do not feel confident with small-group discussions. In particular, evaluation of materials with a specific focus on teaching socio-scientific issues and developing scientific literacy, such as *AS Public Understanding of Science* course (Osborne *et al.*, 2002) and the *Valuable Lessons* project (Levinson and Turner, 2001), raised particular concerns about teacher efficacy in the use of small-group discussions. For example, Osborne *et al*. (2002) comment, “Our view, given the poor quality of much of the teaching involving discussion, is that training is essential. Teachers … need an opportunity to interact with experienced humanities teachers, to observe the strategies they use for fostering and stimulating discussion …” (p75). In a similar vein, Levinson and Turner recommended that, “Teacher training courses should provide prospective science teachers with more opportunities in the area of initiating and managing discussions … practicing teachers should improve and update such skills through CPD [continuing professional development].” (p21).

It is clear from the discussion above that there is considerable interest in the use of small-group discussion in science teaching. Some of this interest has emerged directly from research studies, whilst, in other areas, it draws more loosely on research evidence and take the form of approaches which are being advocated in science teaching, but whose effects have yet to be explored on a more systematic basis. There also appears to be a comparative lack of guidance for teachers. The evidence presented in this paper provides a number of insights into the form such guidance might take.

The origins and purposes of systematic reviews

Systematic reviews of research studies are a comparatively recent development in education, though they are well established in medical research. They have emerged from the international debate over the nature and purpose of educational research, and how it contributes to maximising the effectiveness of educational provision (*e.g*. Hargreaves, 1996 and Hillage *et al*., 1998, in the UK; Shavelson and Towne, 2001, in the USA).

There are several reasons why systematic reviews are being seen as a key strand in educational research. Firstly, there is a growing interest in practical policy-related decision making being linked to evidence in a number of areas, not just in education. Systematic reviews of research literature are seen as having the potential to yield evidence on which policy makers can draw (Davies, 2000; Torgerson and Torgerson, 2001). Secondly, there is a drive towards forging closer links between research, policy and practice (see, for example, Hargreaves, 1996; OECD, 2002; Oakley, 2002). In particular, drawing on research findings in classroom practice is seen as desirable, with teachers being encouraged to engage in what is variously described as 'evidence-based', 'evidence-informed' or 'evidence-enriched' practice (e.g. Millar *et al*. 2006).

In 2000 the Government in the UK funded, via the Department for Education and Skills (DfES), the Evidence for Policy and Practice Initiative (EPPI)-Centre to focus on undertaking systematic reviews of research evidence in key areas of education, and reporting these in a form accessible to a range of different user groups, including teachers, researchers and policy-makers. The advantages and disadvantages of systematic reviews when compared with other forms of review, and of reviews employing the EPPI methodology have been rehearsed elsewhere (Bennett *et al.*, 2005).

The systemic review process, as developed by the EPPI-Centre, involves several stages:

- identification of review topic area and review research question or questions;
- development of *inclusion and exclusion criteria* for studies in the review (relating to, for example, aspects such as the age of students, the nature of the research design, and the reported outcomes);
- undertaking of *systematic searches* of electronic data bases and other sources for potentially relevant research studies;
- refining the search through *screening* the potentially relevant studies against the inclusion criteria;
- coding or *keywording* studies against pre-specified and agreed characteristics (some of which are generic to all EPPI reviews, whilst others are developed specifically for each review);
- production of an overview or *systematic map* of studies in the review area, that groups the studies according to their chief characteristics;

- undertaking an *in-depth review* of studies to look in detail at their design and findings and to evaluate the quality of the evidence reported.

This information is then used to make judgements about the quality of the weight of evidence presented in the study in relation to the review research question. Each of these judgements involves a decision about whether the weight of evidence in a study is *high*, *medium* or *low* Full details of the EPPI methodology may be found in the EPPI Review Group Manual (EPPI-Centre, 2002).

The EPPI review methods tend to be more tailored to quantitative research. As much of the work on small-group discussion draws extensively on qualitative approaches, the work reported here extended the EPPI review methods to draw on the guidance and framework for assessing research evidence in qualitative research studies (Spencer *et al.*, 2003). This was seen as particularly important as the majority of the studies included in the review made use of qualitative methods. As such, they ran the risk of being characterized as low in quality had judgments been informed only by the EPPI criteria.

The systematic review of studies on small-group discussions in science lessons

The main review research question was *How are small-group discussions used in science teaching with students aged 11-18, and what are their effects on students' understanding in science?* Within this, two reviews were conducted. The purpose of the first review was to gain an overview of the nature of small-group discussions used in science lessons in order to assess the levels of use, establish any patterns in use, such as in topic focus, group structure and the dynamics of group interactions. As much of the support for the use of small-group discussions is linked to their potential benefits associated with to improved understanding of science ideas, the second review focused specifically on the effects of small-group discussions on understanding.

A note on terminology

The reports on small-group discussions scrutinised for the systematic reviews made it clear that the term was used in a wide variety of ways. For the purposes of the reviews, a small-group discussion was taken to be an activity that:

- involves groups of two to six students;
- has a specific stimulus (e.g. a newspaper article, video clip, prepared curriculum materials; structured teacher input);
- involves a substantive discussion task of at least two minutes (i.e. did not simply involve a student talking to a neighbour briefly about an idea or to agree and answer to a question);
- is either *synchronous* (i.e. face-to-face) or *asynchronous* (i.e. mainly IT-mediated);
- has a specific purpose (e.g. individual sense-making, or leading to an oral presentation, or to a written product).

The term *understanding* has been taken to encompass understanding of science concepts, understanding of ideas about the nature of science and understanding of the methods of science.

Studies in the review

The reviews focused on research on teaching at high school level, undertaken in the period 1980-2005, and published in English. Student age was restricted to 11-18 because this is the age range covered by the majority of reported studies, pointing to this being the school age range where the use of small-group discussions has been promoted most actively. The start date for the period of publication was selected because this was the time when the use of small-group discussions started to become more prominent in science teaching. The inclusion criteria for studies in the review were: (i) they were about the use of small-group discussions in science lessons, (ii) they involved groups of two to six students, (iii) they focused on a substantive, structured discussion task of two minutes' duration or more, and, for the second review, (iv) they addressed aspects of students' understanding in science.

94 studies were identified that met the above inclusion criteria. The identification of the studies is a two-step process. Firstly, systematic searches are undertaken, principally through the use of electronic search strings. Electronic searching inevitably means that large numbers of studies emerge in the initial stage, and some 2,290 studies matched the search terms. The second step involves refining the search through careful screening of abstracts (or full copies

of reports if there is insufficient information in the abstract) against the inclusion criteria. The 94 studies identified were then coded for particular characteristics (*keyworded*) to produce an overview (*the systematic map*) of studies of small-group discussion interventions. In producing the map, the following characteristics of studies scrutinised included the country of study, the age/level of the students, the type of study, the science discipline of the study, constitution of discussion groups, the duration of discussion tasks, the stimulus provided, organisational features of discussion tasks, the product of discussion tasks, the research strategy used to gather data, the nature of data collected, and the outcomes reported.

An overview (systematic map) of studies on small-group discussions

This section presents a brief overview of the key features to emerge from the 94 research reports on the use and effects of small-group discussions.

Although the period of the review covered 1980-2005, over 90% were from post-1990, indicating that that the research activity area has been minimal up to fifteen years ago and has been most prolific from the late 1990s onwards. The majority of the reported work has been undertaken in North America (USA = 39%, Canada = 12%) and the UK (13%). Other work has been undertaken in Australia, The Netherlands and Germany. However, these figures need to be set in the context of the review being limited to reports published in English, though the review does include reports of studies of small-group discussions held in Bahasa Malay, Cantonese, Dutch, Finnish, French, German, Greek, Hebrew, Mandarin, Portuguese and Spanish.

Relatively little detail was given about how groups were constituted, but where this was provided, mixed-ability and friendship groups predominated. In one-third of cases, groups were deliberately constituted by the teacher, with students choosing their groups in the remaining cases. Slightly over half the studies used a group size of 3-4 students, with a further quarter of studies using pairs of students for groups. Over 80% of the studies concerned self-contained and permanent groups. The remaining studies drew on the techniques of 'snowballing', 'envoying' and 'jigsawing'. Snowballing discussions involve progressively larger groups of students discussing a question or idea and agreeing on their views. Discussion starts with pairs, who then join together and so on. Envoying discussions involve groups of students discussing a common task. When the discussion is completed, one member of each group moves to another group to report on the discussion of the original group and to hear about the discussion in the second group. The envoy then returns to the original group to report back. Jigsawing discussions have students first working in groups where each has a different task and then moving into different groups that comprise of all the people who have focused on the same aspect of the task.

Two-thirds of the discussions took a class period or longer, with approximately two-fifths taking place in lessons described as 'science', a similar proportion in physics lessons, around one-fifth in biology lessons and only 4% in chemistry lessons. One possible explanation for this very small figure for chemistry is that most of the small-group discussions relating to exploration of difficult ideas are located mainly within physics, and those developing skills of decision-making on socio-scientific issues are more commonly placed within biology classes.

The most common stimulus material provided for the discussion tasks was printed worksheets, which were used in over two-thirds of the studies. Practical work and computer software provided the stimulus for around two-fifths and a quarter of discussions respectively. Some discussion made use of more than one stimulus. Rather surprisingly, only one study used a newspaper article as a stimulus for discussion.

In a very high proportion of the studies (94%), the main aim of the discussion task was individual understanding of the science underlying the activity, such as in a practical experiment, the preparation of a poster or a computer-based exercise, in which the learners were engaged. In the majority of cases this understanding was then shared with classmates in different ways: groups might present their findings or views orally (20%) or by way of posters (10%) or might defend their position in a whole class debate (5%). There were surprisingly few examples of written products being generated directly as an outcome of the discussion (6%).

Around three-fifths of the studies reported on evaluations of small-group discussions with the remainder providing descriptive information about the use of small-group discussions. Case studies featured prominently, with extensive use being made of video and audio recordings in order to gather detailed information about the nature of discussions. One outcome of the very labour-intensive nature of much of the data collection and analysis was that sample sizes tended to be small – very often one class or one or two groups of students within a class. Studies involving several classes, or classes in more than one school, were rare.

The chief characteristics of research on small-group discussions are summarised in Table 1.

[Table 1 about here.]

The in-depth reviews

Two in-depth reviews were conducted, covering a total of 24 of the 94 studies in the systematic map. These 24 studies were selected because (i) they reported in detail on the use of small-group discussions in science lessons (19 studies), and/or (ii) they reported evaluations of interventions aimed at developing aspects of students' understanding in science (14 studies).

Studies were rated as high (H), medium high (MH), medium (M), medium low (ML) or low (L). As quality judgements in systematic reviews are made in relation to the specific focus of each of the reviews, some studies were given different ratings in each review. Ratings were based on the extent to which the studies reported met a range of criteria relating to (i) the nature of the sample and how it was selected; (ii) the nature of any control group (for evaluations); (iii) the extent to which small-group discussions formed the main feature or variable being investigated (for evaluations); (iv) the level of detail proved about the discussion task; (v) the steps taken to establish the reliability and validity of the data collection tools and processes, and the data analysis; (vi) the trustworthiness and relatability of the data collection and analysis for qualitative data; (vii) the representativeness of the data collection situation to normal classroom situations; (viii) the quality of the reporting. No studies met all the relevant criteria, so none were judged to be high quality. The studies rated as 'medium high' met most of the relevant criteria. Studies rated at the lower end of the scale displayed one or more of the following characteristics: they provided comparatively little detail of the discussion tasks, they tended to be weaker in the reporting of steps taken to enhance the reliability and validity of data collection and analysis, and/or they were overly-descriptive at the expense of analysis and discussion. The categorisation process is described in detail in the full technical reports (see References section).

Table 2 lists the studies and their quality rating for each review.

[Table 2 about here.]

The remaining discussion focuses on the nineteen studies rated as 'medium' or better, as these provide the stronger evidence, and Table 3 summarises the key features of each of these studies.

[Table 3 about here.]

The detailed evidence from the in-depth reviews

The consolidated evidence presented draws primarily on the findings of the fourteen studies in the Review 1 and the twelve studies in Review 2 weighted as medium or better in overall quality, as these studies generated the better quality evidence. Of these studies, nine were common to both reviews, i.e. they focused on aspects of the use of small-group discussions and the development of understanding. The features discussed below are considered to be those that offer the strongest evidence as they emerged from three or more studies. To avoid undue repetition, evidence from both reviews is considered together under the following five headings: focus of the discussion topic, group structures and interactions, negotiating and agreeing meaning through discussion, effects on understanding, factors promoting effective discussion to enhance understanding

Focus of the discussion topic

The studies were based on a range of science topics, as Table 3 demonstrates. Seven studies addressed aspects of understanding of science topics. These focused on light (Roth and Roychoudhury, 1992), kinetic theory (Palincsar *et al.*, 1993), genetics (Finkel, 1996; Jiménez-Aleixandre *et al.*, 2000), two physics topics [shadows, floating and sinking] (Woodruff and Meyer, 1997), mechanics (Tao, 2001), and density (Kurth *et al.*, 2002). Four looked primarily at aspects of what could be termed scientific method: hypotheses on the diagnosis of disease (Richmond and Striley, 1996; Lajoie, 2001), designing controlled experiments (Sherman and Klein, 1995), and building theories and models from primary evidence on elements and bonding (Keys, 1997). Three had a specific focus on socio-scientific issues in relation to the greenhouse effect (Gayford, 1995), genetic engineering (Zohar and Nemet, 2002), and environmental science (Jiménez-Aleixandre and Pereiro-Muñoz (2002). Three involved making predictions based on evidence presented in the topics of sound (De Vries *et al.*, 2002), a range of biology topics (Lavoie, 1999), and mechanics (Tolmie and Howe, 1993). The studies by Hogan (1999a and 1999b) had no specific topic focus but were based on a series of discussion task developed by the researchers.

The studies in the in-depth review reflect the patterns noted in the systematic map, where the bulk of the discussion topics lay in the areas of physics and biology. Although there was diversity in topic focus, the common link between the discussions focusing on the development of understanding was that they all required students to draw on evidence to support a particular hypothesis, theory or point of view.

Group structures and interactions

The principal evidence on group structure and interaction came from five studies: Tolmie and Howe (1993), and Richmond and Striley (1996); Keys (1997); Hogan (1999a); De Vries *et al.*, 2002.

Group leadership emerged as crucial in promoting effective discussion. Richmond and Striley (1996) and Kurth et al. (2002) established the need for a leader to adopt an inclusive style and share tasks equitably around a group, as this promoted more substantial engagement in the discussion by a number of participants, and increased the quality of the discussion. This, in turn, permitted most members to develop their understanding. Non-inclusive leadership generated much off-task talk and engagement was generally low. Hogan (1999a) found that at least one group member had to act in a way which promoted reflection in the group for understanding of science ideas to be developed.

The evidence on assigning specific roles to students was mixed. Kurth *et al.* (2002) advocated assigning particular roles to pupils in groups as a means of achieving effective discussion. However, Richmond and Striley, (1996) and (Hogan, 1999a) reported allocating roles to have benefits when tasks were well-structured but counterproductive in poorly-structured tasks, adding to students' difficulties in engaging with the task. Hogan identified eight sociocognitive roles in group reasoning processes which took place in small-group discussions, all of which were persistent over time, and independent of particular formal roles that might have been allocated to students in advance of a discussion. Four of Hogan's roles were positive (promoter of reflection, contributor of content knowledge, creative model builder, mediator of social interaction and ideas) and four were negative (promoter of distraction, of acrimony, of simple task completion, reticent participant).

Aspects of group composition were explored in four studies (Tolmie and Howe, 1993; Keys, 1997; Hogan, 1999a; De Vries *et al.*, 2002). Tolmie and Howe found improved understanding to be independent of group composition (male, female, mixed), with biggest improvements being noted when groups contained members with a high degree of dissimilarity in their initial predictions and explanations. In common with Keys and De Vries *et al.,* Tolmie and Howe identified clear differences in interactional styles with all-male groups confronting differences in their individual predictions and explanations, whilst all-female groups searched for common features of their predictions and explanations in order to avoid conflict. Mixed groups interacted in a more constrained way than single-gender groups, though they also tended to avoid conflict and look for common patterns in contributions. Tolmie and Howe suggested that both male pairs and female groups demonstrated qualities one would want to see in the development of arguments but did not see this as a reason for promoting the use of mixed gender groups, as their study suggested the best of the all-male and all-female group interactions was lost in mixed pairs. Hogan also found that friendship groups, which were generally single-sex, functioned more effectively and promoted better development of understanding than mixed or teacher-constituted groups.

The studies suggest that both the behavioural characteristics and gender composition of groups need careful consideration if groups are to function purposefully during discussion tasks.

Negotiating and agreeing meaning through discussion

Three of the studies (Roth and Roychoudhury, 1992; Keys, 1997; Jiménez-Aleixandre *et al.*, 2000) reported in detail on the ways in which meanings were negotiated and agreed by groups.

Keys (1997) identified three common characteristics of reasoning in discussions: recognising that prior ideas (models) may be incorrect; evaluating new observations for consistency with current ideas and using evidence to modify ideas; and coordinating all mutually consistent knowledge propositions into a coherent model. A similar pattern was described by Roth and Roychoudhury (1992) who found discussions usually involved positions being stated and contested, with views either accepted or temporarily or permanently rejected as positions finally stabilised into shared meaning. Less positively, they found that students tended not to engage very often in processes which fostered meaning. Rather they would reach agreement on the basis of finding something agreeable to all group members. Agreements were often reached by one or more group members exerting authority, and on the basis of 'majority rule', rather than agreed shared understanding. These findings were echoed by Jiménez-Aleixandre *et al.* (2000) who reported that a large proportion of student talk related to what they termed 'doing the lesson', or interactions referring to the rules of the task, rather than talk related to the focus of the task. Jiménez-Aleixandre *et al.* also noted that arguments were frequently developed by a subset within the group and, though agreement was generally reached, this was often for social reasons, rather than because of agreement over the outcome of the discussion.

Whilst the above findings suggest that groups can work together to develop common understandings, they also suggest that a number of factors can influence the way in which these understandings are reached, and relative lack of engagement with the science content, coupled with the influence of students prepared to express views very strongly means that the gains in understanding may not be that high.

Effects on understanding

All the studies that looked at development of understanding reported benefits from small-group discussion work, though this finding does need to be interpreted with some caution as the majority of those undertaking the research were advocates of the approach. The following provide examples of the effects that were reported.

Roth and Roychoudhury (1992) established that group discussion over the construction of a concept map provided a vehicle for negotiation of meaning and understanding of concepts and their relationships, thus providing a structure through which students were able to learn the language patterns of science and use these to construct scientific knowledge. Tolmie and Howe (1993) reported significant improvements in students' predictions of the trajectories of falling objects through the use of a computer-based simulation. Richmond and Striley (1996) noted increasing levels of sophistication and increased use of subject knowledge in the arguments students developed in discussion of socio-scientific issues. Similarly, Zohar and Nemet (2002) reported substantial changes in the quality of student arguments in the context heredity and genetics.

Factors promoting effective discussion to enhance understanding

The factors that emerged that contributed to effective discussions and enhancement in understanding are of particular interest. Findings from four studies pointed to improvements in understanding being greatest for discussion tasks where there was dissimilarity or conflict in understanding or views. This might take the form of either *internal conflict*, or differences held by individual group members (Tolmie and Howe, 1993; De Vries *et al.*, 2002), or *external conflict* where an external stimulus presents a group with conflicting views (Tolmie and Howe, 1993; Gayford, 1995; Finkel, 1996). In some of the studies the discussion topic was selected to provide opportunities for both internal and external conflict. For example, Tolmie and Howe (1993) required students to make individual predictions about aspects of forces and motion, then engage in a task which required a joint prediction (internal conflict) and finally to compare this with an actual situation to reach an explanation of any discrepancies (external conflict). Whilst other studies did not comment specifically on the need for dissimilarity, it was clear from some of the accounts (e.g. Zohar and Nemet, 2002) that internal and external conflict were built into the discussion tasks. It seems likely that the dissimilarity in views provides a very clear and immediate focus to engage students in discussion.

Two studies offered comments on the nature of the data provided to students for the discussion. Jiménez-Aleixandre *et al.* (2000) indicated that hypothetical, unquestionable data (provided by the teacher) generates different patterns of discussion than empirical, uncertain data, perhaps gathered by students themselves, with the former leading to greater gains in understanding. In a similar vein, Roth and Roychoudhury (1992) found discussion to be more productive if students were provided with a fixed set of concepts to delimit the content of the discourse.

Three studies pointed to improved understanding when students were given specific instructions on how to construct arguments or cues to guide them in the points they needed to include (Sherman and Klein, 1995; De Vries *et al.*, 2002; Zohar and Nemet, 2002). This finding was reflected in the more general observations in two further studies that scaffolding routines, or structuring discussion through the provision of interim targets, also improved students' understanding (Palincsar *et al.*, 1993; Finkel, 1996).

Although gains in understanding were reported, the studies also suggested that students often struggled to formulate and express coherent views during small-group discussions, and demonstrated a low level of engagement with tasks. It is therefore not surprising that a number of the studies made recommendations relating to the need for students and/or teachers to be given explicit teaching in the skills associated with the development of arguments and the characteristics associated with effective group discussions.

Three studies (Richmond and Striley, 1996; Hogan, 1999a; Zohar and Nemet, 2002) recommended training for both students and teachers in the skills needed for handling and participating in group discussions. Richmond and Striley indicated that productive learning was unlikely to take place on a large scale through the use of small-group discussions until students acquire the skills associated with inclusive leadership and are thus able to foster a climate of equitable participation. Hogan (1999a) argued that guiding students towards taking constructive roles in discussions could be achieved through metacognitive training, i.e. knowledge about the nature of collaborative learning, effective group learning strategies, and awareness of what constitutes progress. Two studies (Jiménez-Aleixandre *et al.*, 2000 and Roth and Roychoudhury, 1992) recommended coaching in argumentation skills for both teachers and students.

Jiménez-Aleixandre *et al.* (2000) suggested that effective discussions are only likely to take place when linked to specific, inquiry-focused tasks where help is given to students to develop their understanding through the construction of arguments. Similarly, Roth and Roychoudhury (1992) reported that students frequently struggled with language, often making short utterances, and appeared to find it difficult to clarify their understanding through explanations, justifications and elaborations. This led them to conclude that a major outcome of their study was the recognition that understanding was only likely to be improved is students were given help in constructing arguments.

One study (Zohar and Nemet, 2002) did involve incorporating explicit instruction about argumentation into their intervention. One introductory lesson involved arguments being defined and their structure explained, together with providing examples of characteristics of good arguments. Students then practised the principles through several concrete examples. Zohar and Nemet concluded that argumentation skills were enhanced by explicit instruction and several opportunities for students to take part in discussions to help develop their skills.

The nature of the evidence presented in a substantial number of the studies points to the importance of training for teachers and students. This is all the more important as many of the researchers involved in the studies were committed to the use of small-group discussions and had developed some proficiency in designing discussion tasks. This review finding also resonates with the recommendations of Levinson and Turner (2001) and Osborne *et al*. (2002) in their evaluations of two programmes focusing on teaching of socio-scientific issues, and noted earlier in this paper.

Methodological considerations

Although the primary focus of the reviews was to gather substantive data on the use of small-group discussions, some of the methodological aspects have a bearing on the nature of evidence yielded.

Positive features of the data collection included the use of multiple data sources with all studies drawing on at least two different kinds of data to increase trustworthiness. Whilst virtually all studies used video recording and/or audio recording to make verbatim records of discussions, these were supported by direct observation to record field notes, interviews, products of student tasks, such as concept maps, student questionnaires and measures of student knowledge were obtained. Although the methods used were rarely justified, the picture gained was one of studies collecting extensive data in an attempt to get as detailed a picture as possible of students' dialogue.

There were a number of limitations to the data collection. All the studies used a convenience sample for the identification of schools, in many cases using schools where access had been secured through previous involvement of the researcher. With one exception (Zohar and Nemet, 2002), the studies were based in one school and often within one class. A characteristic of much of the work was the use of retrospective sampling, i.e. data were gathered on a number of groups, but reports presented on only a sample of the groups within this, depending on characteristics of the discussion which emerged in the analysis. Such sampling methods are probably realistic for research studies fitting in with practice, and requiring extensive periods of data collection and thus a high degree of co-operation with the class teachers involved. However, retrospective sampling does confer the option on the researcher of exercising a high degree of selectivity in relation to the data presented.

Three sizeable studies (Gayford, 1995; Lavoie, 1999; Zohar and Nemet, 2002) utilised an experimental design, making comparisons between a control group who experienced a more conventional teaching approach with a group which received some form of intervention related to small-group discussion work. One study (Tolmie and Howe, 1993) specifically set up groups where gender was a variable to be explored. However, the emphasis of the majority of the studies was on describing and interpreting the nature of student discussions and their effects on students' understanding, sometimes making detailed comparisons between groups participating in their studies. Two factors may contribute to the absence of a control group in the studies. Firstly, those undertaking the research might see no need to design their studies to include a control group in what were largely qualitative and interpretative studies. Secondly, the practicalities associated with collecting and analysing extensive in-depth data from a much larger sample in order to make such comparisons would place prohibitive resource constraints on the studies.

Data analysis was characterised positively by the presentation and discussion of rich and detailed data in the form of extracts from students' discourse, with all studies adopting procedures to increase the trustworthiness of the analysis by having two or more people involved. However, given that the studies were largely gathering qualitative data, there was a surprising lack of contextual detail. Data also tended to be presented in a rather convergent manner, with few examples of data being presented which might disprove assertions or report on unintended outcomes.

Of particular interest were the two very contrasting overall analysis strategies apparent in the studies. The first strategy, adopted by the majority of the studies, was to develop grounded theory from the data through the development of categories then used to characterise the interactions between participants. Hogan (1999a) referred to her analysis as ‘ethnographic interaction analysis’, whilst Roth and Roychoudhury (1992) used what they described as the techniques used by anthropologists when analysing interactive behaviours. The second strategy, whose use was more limited, was to draw on existing work on discourse analysis or discourse analysis classifications. Such an approach involves trying to identify themes in what people say by looking at sentences, groups of sentences or sentence fragments. These might, for example, relate to attempts to cite others to support a view, or use of evidence to support an account of an event. Discourse analysis techniques were employed by Jiménez-Aleixandre *et al.* (2000), who drew on the work of Bloome *et al.* (1989) to do the initial coding of exchanges between students, and then used Toulmin’s (1958) work on argument to classify the interactions where students were talking about science aspects in the discussion. Keys (1997) drew on elements of a framework developed by Kuhn (1993) to code students’ verbal interactions relating to scientific reasoning. However, there was a notable absence of justification in the studies for the approach adopted for analysis, with the development of grounded theory appearing to be seen as an unproblematic choice in the majority of cases. It may be the case that the choice of approach reflects the personal views of the researchers on the role and purposes of data in research. However, the lack of reference made to discourse analysis techniques suggests that these approaches may be unfamiliar to some researchers working in the area of small-group discussions, which, in turn, may be limiting the nature of the analysis. There would appear to be a good case for those researching the effects of small-group discussions to gain a greater familiarity with discourse analysis techniques.

Table 4 summarises the key findings to emerge from studies on the use and effects of small-group discussions in science.

[Table 4 about here.]

Conclusions and recommendations

The reviews reported here have yielded insights on both the substantive focus (small-group discussions and their effects) and on the methods employed to gather the data. The review has revealed a number of features of particular interest in relation to the use of small-group discussion work in science. It is clear from the review that a complex and interacting set of factors are involved in enabling students to engage in dialogue in a way that could help them draw on evidence to develop and articulate their understanding of science ideas.

Two particularly strong features which have emerged from the work undertaken for the review is that there is a relatively little good quality systematic research on the effects of small-group discussion work, and considerable uncertainty on the part of teachers as to what they are required to do to implement good practice. Current policy is strongly advocating the use of small-group discussion work in science, and the reviews do indicate that there could be benefits arising from this, as small-group discussion work can provide an appropriate vehicle for assisting in the development of students' understanding of science ideas. Thus teachers should be encouraged to incorporate such discussions into their teaching. However, it is also clear that small-group discussion work needs to be supported by the provision of support and guidance for teachers and students on the development of the skills necessary to make such work effective.

One feature, notable by its absence, was the dearth of systematic evidence on the effects of the use of small-group discussions on students' attitudes to their science lessons or science more widely. The absence of such data was very surprising, as the motivational effects of small-group discussions are often cited as a reason for their inclusion in science teaching.

It is clear from this review that there is considerable variation in the nature of research into small-group discussion work, particularly in relation to its focus, the clarity with which any variables being investigated are specified, the use of opportunistic samples for data collection, and the techniques used to analyse data. Particularly striking are the two very contrasting approaches to data analysis, with some studies developing grounded theory from the data, and others drawing on existing models to structure their analysis. A substantial proportion of the work also focuses on descriptive data. This can be very helpful in the early stages of a new research area. However, with increasing interest in the *effects* of small group discussions – on student learning, understanding, and attitudes – there is a need to consider what strategies and techniques lend themselves best to the gathering and analysis of data that would help explore such effects.

Taken together, these findings suggest there are four key features that need to characterise any further developments in the use of small-group discussions in science teaching and research into their effects. Firstly, some form of professional development training for teachers is highly desirable to provide guidance on how to maximise the effectiveness of small-group discussions. Secondly, further research into the effects of small-group discussions should include a consideration of the extent to which analysis of the data might benefit from established discourse analysis techniques developed in other subject areas (e.g. Barnes and Todd, 1997; Mercer and Littleton, 2007), to establish what they might have to offer work in science. Thirdly, the area would benefit from a more detailed exploration of the effects of small-group discussions on attitudinal effects. Finally, in relation to providing evidence of the *effects* of small-group discussions, there would appear to be potential benefits associated with adopting a mixed method approach to data collection, marrying in-depth qualitative data on the nature of discussions with more quantitative data on student attributes.

Acknowledgements

The work reported in this paper was funded by the Department for Education and Skills (DfES) as part of the *Evidence, Policy and Practice Initiative* (EPPI).

Electronic searches conducted for the review

Electronic searches of the following databases were conducted:

British Education Index (BEI)
Educational Resources Information Center (ERIC)
PsycINFO
Social Science Citation Index (SSCI)

References in text of report

Barnes, D. & Todd, F. (1997) Communication and learning in small groups. London: Routledge and Kegan Paul.

Bennett, J., Lubben, F., Hogarth, S. & Campbell, B. (2005) Systematic reviews of research in science education: rigour or rigidity? *International Journal of Science Education,* 27 (4), 387-406.

Bentley, D & Watts, M. (1992) Communicating in school science. London: Falmer Press.

Black, P. & Wiliam, D. (1998) *Inside the black box: raising standards through classroom assessment.* London: King's College.

Davies, P. (2000) The relevance of systematic reviews to educational policy and practice. *Oxford Review of Education*, 26, 365-378.

Daws, N. & Singh, B. (1996) Formative assessment: to what extent is its potential to enhance pupils' science being realised? *School Science Review*, 77 (281), 93-100.

Department for Education and Employment (DfEE) (1998) *The National Literacy Strategy*. London: DfEE.

Driver, R. & Bell, B. (1985) Students' thinking and the learning of science: a constructivist view. *School Science Review*, 67 (240), 443-456.

Driver, R., Asoko, H., Leach, J., Mortimer, E. & Scott, P. (1994) Constructing scientific knowledge in the classroom. *Educational Researcher*, 23 (7), 5-12.

EPPI-Centre (2002) *EPPI-Centre Review Group Manual*. London: Social Science Research Unit, Institute of Education.

Hargreaves, D. (1996) *Teaching as a research-based profession: possibilities and prospects*. London: The Teacher Training Agency (TTA).

Hillage, L., Pearson, R., Anderson, A. & Tamkin, P. (1998) *Excellence in research on schools.* Brighton: Institute for Employment Studies.

Kuhn, D. (1993) Science as argument: implications for teaching and learning scientific thinking. *Science Education,* 77 (2), 319-337.

Kyriacou, C. (1998) Essential teaching Skills (2nd edition) Cheltenham: Stanley Thornes.

Lemke, J. (1990) Talking science. New York: Ablex Publishing Corporation.

Levinson, R. & Turner, S. (2001) *Valuable Lessons: Engaging with the Social Context of Science in Schools.* London: The Wellcome Trust.

Mercer, N. & Littleton, K. (2007) Dialogue and the Development of children's thinking: a sociocultural approach. Oxford: Routledge.

Millar, R. (2006). *Twenty First Century Science*: Insights from the design and implementation of a scientific literacy approach in school science. *International Journal of Science Education*, 28 (13), 1499-1521.

Millar, R. & Osborne, J. (eds) (1998) *Beyond 2000: Science Education for the Future*. London: King's College/The Nuffield Foundation.

Millar, R., Leach, J., Osborne, J. & Ratcliffe, M. (2006) *Improving subject teaching: lessons from research in science education.* London: Routledge.

Newton, P., Driver, R. & Osborne, J. (1999) The place of argumentation in the pedagogy of school science. *International Journal of Science Education,* 21 (5), 553-576.

Oakley, A. (2002) Social science and evidence-based everything: the case of education. *Educational Review*, 54(3), 21-33.

OECD (2002) *Educational research and development in England: examiners' report.* Organisation for Economic Co-operation and Development. www.oecd.org/dataoecd/17/56 (accessed 2 January 2008)

Osborne, J. & Collins, S. (2001) Pupils' views of the role and value of the science curriculum. *International Journal of Science Education*, 23 (5), 441-467.

Osborne, J., Duschl, R. & Fairbrother, R. (2002) *Breaking the Mould? Teaching Science for Public Understanding*. London: The Nuffield Foundation.

Scott, P. (1998) Teacher talk and meaning-making in science classrooms: a Vygotskyan analysis and review. Studies in Science Education, 32 45-80.

Shavelson, R. & Towne, L. (2001) (eds) *Scientific enquiry in education*. Washington DC: National Academy Press.

Sinclair, J. and Coulthard, R. (1975) *Towards an Analysis of Discourse,* OUP.

Spencer, L., Ritchie, J., Lewis, J. & Dillon, L. (2003). *Quality in Qualitative Evaluation: a framework for assessing research evidence*. London: The Strategy Unit.

Sutton, C. (1992) Words, science and learning. Buckingham: Open University Press.

Toulmin, S. (1958). *The use of argument*. Cambridge: Cambridge University Press.

Torgerson, C. (2003) *Systematic Reviews*. London: Continuum.

Torgerson, C. & Torgerson, D. (2001) The need for randomised controlled trials in educational research. *British Journal of Educational Studies*, 49 (3), 316-328.

Wellington, J., & Osborne, J. (2001). *Language and literacy in science education*. Buckingham: Open University Press.

References to the full systematic review technical reports

Author 1, Author 3, Author 2, Author 4, & Author 5. (2004) A systematic review of the nature of small-group discussions in science teaching aimed at improving students' understanding of evidence. In: *Research Evidence in Education Library*. London: EPPI-

Centre, Social Science Research Unit, Institute of Education. http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=321

Author 1, Author 3, Author 2 & Author 4. (2004) A systematic review of the use of small-group discussions in science teaching with students aged 11-18, and their effects on students' understanding in science or attitude to science. In: *Research Evidence in Education Library*. London: EPPI-Centre, Social Science Research Unit, Institute of Education. http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=326

References for the studies included in the review

De Vries E, Lund K, Baker M (2002) Computer-mediated epistemic dialogue: explanation and argumentation as vehicles for understanding scientific notions. *Journal of the Learning Sciences* 11: 63-103.

Finkel EA (1996) Making sense of genetics: students' knowledge use during problem solving in a high school genetics class. *Journal of Research in Science Teaching* 33: 345-368.

Gayford C (1995) Science education and sustainability: a case-study in discussion-based learning. *Research in Science & Technological Education* 13: 135-45.

Hogan K (1999a) Sociocognitive roles in science group discourse. *International Journal of Science Education* 21: 855-882.

Hogan K (1999b) Thinking aloud together: a test of an intervention to foster students' collaborative scientific reasoning. *Journal of Research in Science Teaching* 36: 1085-1109.

Jiménez-Aleixandre MP, Pereiro-Muñoz C (2002) Knowledge producers or knowledge consumers? Argumentation and decision making about environmental management. *International Journal of Science Education* 24: 1171-1190.

Jiménez-Aleixandre MP, Rodriguez AB, Duschl RA (2000a) 'Doing the lesson' or 'doing science': argument in high school genetics. *Science and Education* 84: 757-92.

Johnson SK, Stewart J (2002) Revising and assessing explanatory models in a high school genetics class: a comparison of unsuccessful and successful performance. *Science and Education* 86: 463-480.

Keys CW (1997) An investigation of the relationship between scientific reasoning, conceptual knowledge and model formulation in a naturalistic setting. *International Journal of Science Education* 19: 957-970.

Kurth LA, Anderson CW, Palincsar AS (2002) The case of Carla: dilemmas of helping all students to understand science. *Science Education* 86: 287-313.

Lajoie SP, Lavigne NC, Guerrera C, Munsie SD (2001) Constructing knowledge in the context of BioWorld. *Instructional Science* 29: 155-186.

Lavoie DR (1999) Effects of emphasizing hypothetico-predictive reasoning within the science learning cycle on high school student's process skills and conceptual understandings in biology. *Journal of Research in Science Teaching* 36: 1127-1147.

Meyer K, Woodruff E (1997) Consensually driven explanation in science teaching. *Science Education* 81: 173-192.

Palincsar AS, Anderson C, David YM (1993) Pursuing scientific literacy in the middle grades through collaborative problem solving. *Elementary School Journal* 93: 643-658.

Richmond G, Striley J (1996) Making meaning in classrooms: social processes in small-group discourse and scientific knowledge building. *Journal of Research in Science Teaching* 33: 839-858.

Roth W-M, Roychoudhury A (1992) The social construction of scientific concepts or the concept map as conscription device and tool for social thinking in high school science. *Science and Education* 76: 531-557.

Sherman GP, Klein JD (1995) The effects of cued interaction and ability grouping during cooperative computer-based science instruction. *Educational Technology Research and Development* 43: 5-24.

Suthers D, Weiner A (1995) Groupware for developing critical discussion skills. In: Schnase JL, Cunnius EL *Proceedings of CSCL '95: The First International Conference on Computer Support for Collaborative Learning*. New Jersey: Lawrence Erlbaum Associates Inc, pages 341-348.

Tao PK (2001) Developing understanding through confronting varying views: the case of solving qualitative physics problems. *International Journal of Science Education* 23: 1201-1218.

Tao PK (2003) Eliciting and developing junior secondary students' understanding of the nature of science through a peer collaboration instruction in science stories. *International Journal of Science Education* 25: 147-171.

Tolmie A, Howe C (1993) Gender and dialogue in secondary school physics. *Gender and Education* 5: 191-209.

Tsai C-C (1999) 'Laboratory exercises help me memorize the scientific truths': a study of eighth graders' scientific epistemological views and learning in laboratory activities. *Science and Education* 83: 654-674.

Williams A (1995) Long-distance collaboration: a case study of science teaching and learning. In: Spiegel SA *Perspectives from Teachers' Classrooms. Action Research. Science FEAT (Science for Early Adolescence Teachers).*

Woodruff E, Meyer K (1997) Explanations from intra- and inter-group discourse: students building knowledge in the science classroom. *Research in Science Education* 27: 25-39.

Zohar A, Nemet F (2002) Fostering students' knowledge and argumentation skills through dilemmas in human genetics. *Journal of Research in Science Teaching* 39: 35-62.

Talking science: the research evidence on the use of small-group discussions in science teaching

Abstract

This paper reports the findings of two systematic reviews of the use and effects of small-group discussions in high school science teaching. 94 studies were included in an overview (systematic map) of work in the area, and 24 studies formed the basis of the in-depth reviews.

The reviews indicate that there is considerable diversity in the topics used to promote small-group discussions. They also demonstrate that students often struggle to formulate and express coherent arguments, and demonstrate a low level of engagement with tasks. The reviews suggest that groups function more purposefully, and understanding improves most, when specifically constituted such that differing views are represented, when some form of training is provided for students on effective group work, and when help in structuring discussions is provided in the form of 'cues'. Single sex groups function more purposefully than mixed sex groups, though improvements in understanding are independent of gender composition of groups. Finally, the reviews demonstrate very clearly that, for small-group discussions to be effective, teachers and students need to be given explicit teaching in the skills associated with the development of arguments and the characteristics associated with effective group discussions.

In addition to the substantive findings, the paper also reports on key features of the methods employed to gather and analyse data. Of particular note are the two contrasting approaches to data analysis, one adopting a grounded theory approach and the other drawing on established methods of discourse analysis.

Introduction

The use of small-group discussions in teaching has been advocated for a number of years in science teaching, both to motivate students and to enhance their learning. Such tasks are now appearing with greater frequency in a range of high school science teaching resources. Despite this increasing popularity, comparatively little is known about the detail of their use and effects. This paper examines the research evidence on the use of small-group discussions through presenting the findings of two systematic reviews, with a view to making recommendations for policy and practice on the use of small-group discussions in science teaching.

As originally conceived, the reviews had two principal aims. The first of these was to identify the ways in which small-group discussions are currently used in science lessons. The second was to look at the effects of small group discussions on students' understanding of science and on students' attitudes to science. In practice, the review work established that there is a dearth of studies reporting in any detailed and systematic way on the effects of small-group discussions on students' attitudes to science, so this paper reports the review findings on the use of small-group discussions in science teaching and their effects on students' understanding of science ideas.

Several factors have contributed to the current high levels of interest in small-group discussion work in science. The use of small-group discussion in teaching has its origins in learner-centred teaching approaches, and is one of a range of 'active learning' strategies aimed at stimulating students' interest in what they are studying by providing them with a significant degree of autonomy over the learning activity (e.g. Bentley and Watts, 1992; Kyriacou, 1998). A number of people have advocated the use of discussions in science lessons. Lemke (1990) argues that "learning science means learning to talk science" (p1), and that this means moving away from science lessons dominated by teacher talk in which the teacher asks a question, then invites and evaluates a student response. Lemke refers to this as 'triadic dialogue', which is similar to the initiation-response-feedback (IRF) sequence characterised earlier by Sinclair and Coulthard (1975). Lemke and others (e.g. Sutton, 1992, and Wellington and Osborne, 2001) criticise the approach for leading to talk that focuses on what the teacher wants to hear, rather than promoting genuine communication, and all argue for increased use of discussion work in science lessons.

Support for the use of small-group discussions in science teaching has also emerged in the recommendations from other areas of work in science education. For example, research on alternative conceptions has explored in depth the ideas and understanding students bring with them to science lessons and the ways in which some of their ideas may hinder the development of accepted scientific ideas (e.g. Driver *et al.*, 1985). Small-group discussions have been suggested as a means of helping students explore their ideas and move from

Deleted: the constructivist view of learning understandings that may often be naïve to towards more valid scientific ideas and explanations. Further impetus for the inclusion of small-group discussions in science lessons has come from the development of ideas about *social constructivism* (Driver *et al.*, 1994), which draws on the work of Vygotsky and the importance to learning of social dynamics of interaction (Scott, 1998).

The publication of *Beyond 2000* (Millar and Osborne, 1998) in the UK stimulated discussion and debate over the nature of the school science curriculum and, in particular, the ways in which it might foster the development of *scientific literacy*. This term embraces the knowledge, understanding and skills young people need to develop in order to think and act appropriately on scientific matters which may affect their lives and the lives of other members of the local, national and global communities of which they are a part. Many of the materials developed include small-group discussions in the repertoire of activities employed in science lessons in order to encourage students to participate in informed discussion and debate of scientific issues (e.g. Millar, 2006). Linked to the development of scientific literacy is that of *ideas about evidence* (Millar and Osborne, 1998), which involves encouraging students to evaluate, interpret and analyse evidence from primary and secondary sources in science, including stories about how important science ideas were first developed, and then established and finally accepted. This has led to considerations of the role of *argument* in school science, in the sense of putting forward claims and supporting them with sound and persuasive evidence (e.g. Newton *et al.*, 1999; Osborne *et al.*, 2001). Small-group discussions are viewed as having a key role to play here, since the practice of using evidence in argumentation requires interaction with peers.

Formatted: Right: -18 pt

Deleted: There are now moves in a number of countries to introduce science curricula with a strong emphasis on scientific literacy.

Deleted: .

Deleted: ,

Formatted: Font: (Default) Arial, 11 pt

In the UK, two other developments have also served to raise the profile of small-group discussions. Firstly, the publication of *Inside the Black Box* (Black and Wiliam, 1998) has resulted in considerable attention being paid to formative assessment, or assessment for learning. Small-group discussions are one approach that has been advocated for increasing the use and effectiveness of formative assessment in science teaching (see, for example, Daws and Singh, 1999). Secondly, there is a more general drive to improve students' *literacy skills*. In England and Wales, this has been formalised into the National Literacy Strategy (DfEE, 1998), where small-group discussions have been advocated as a means for developing students' oral communication skills in science.

Deleted: . S

Deleted: language skills

Deleted: (e.g. Newton *et al.*, 1999, and Osborne *et al.*, 2001).

Whilst the preceding discussion point to a general recognition of the potential value of small-group discussion in science teaching, their use challenges the established pedagogy of science teaching and place new demands on teachers (see, for example, Bentley and Watts, 1992). There is a growing body of evidence, both anecdotal and more systematic, that many teachers lack skills and do not feel confident with small-group discussions. In particular, evaluation of materials with a specific focus on teaching socio-scientific issues and developing scientific literacy, such as *AS Public Understanding of Science* course (Osborne *et al.*, 2002) and the *Valuable Lessons* project (Levinson and Turner, 2001), raised particular concerns about teacher efficacy in the use of small-group discussions. For example, Osborne *et al*. (2002) comment, “Our view, given the poor quality of much of the teaching involving discussion, is that training is essential. Teachers … need an opportunity to interact with experienced humanities teachers, to observe the strategies they use for fostering and stimulating discussion …” (p75). In a similar vein, Levinson and Turner recommended that, “Teacher training courses should provide prospective science teachers with more opportunities in the area of initiating and managing discussions … practicing teachers should improve and update such skills through CPD [continuing professional development].” (p21).

It is clear from the discussion above that there is considerable interest in the use of small-group discussion in science teaching. Some of this interest has emerged directly from research studies, whilst, in other areas, it draws more loosely on research evidence and take the form of approaches which are being advocated in science teaching, but whose effects have yet to be explored on a more systematic basis. There also appears to be a comparative lack of guidance for teachers. The evidence presented in this paper provides a number of insights into the form such guidance might take.

Deleted: widespread

Deleted: ¶
A note on terminology¶
¶
The reports on small-group discussions scrutinised for the systematic reviews made it clear that the term was used in a wide variety of ways. For the purposes of the reviews, a small-group discussion was taken to be an activity that: ¶
<#>involves groups of two to six students;¶
<#>has a specific stimulus (e.g. a newspaper article, video clip, prepared curriculum materials; structured teacher input);¶
<#>involves a substantive discussion task of at least two minutes;¶
<#>is either *synchronous* (i.e. face-to-face) or *asynchronous* (i.e. mainly IT-mediated);¶
<#>has a specific purpose (e.g. individual sense-making, or leading to an oral presentation, or to a written product).¶
¶
The term *understanding* has been taken to encompass understanding of science concepts, understanding of ideas about the nature of science and understanding of the methods of science.¶

The origins and purposes of systematic reviews

Systematic reviews of research studies are a comparatively recent development in education, though they are well established in medical research. They have emerged from the international debate over the nature and purpose of educational research, and how it contributes to maximising the effectiveness of educational provision (*e.g*. Hargreaves, 1996 and Hillage *et al*., 1998, in the UK; Shavelson and Towne, 2001, in the USA).

There are several reasons why systematic reviews are being seen as a key strand in educational research. Firstly, there is a growing interest in practical policy-related decision making being linked to evidence in a number of areas, not just in education. Systematic reviews of research literature are seen as having the potential to yield evidence on which policy makers can draw (Davies, 2000; Torgerson and Torgerson, 2001). Secondly, there is a drive towards forging closer links between research, policy and practice (see, for example, Hargreaves, 1996; OECD, 2002; Oakley, 2002). In particular, drawing on research findings in classroom practice is seen as desirable, with teachers being encouraged to engage in what is variously described as 'evidence-based', 'evidence-informed' or 'evidence-enriched' practice (e.g. Millar *et al.* 2006).

Formatted: Font: Italic

In 2000 the Government in the UK funded, via the Department for Education and Skills (DfES), the Evidence for Policy and Practice Initiative (EPPI)-Centre to focus on undertaking systematic reviews of research evidence in key areas of education, and reporting these in a form accessible to a range of different user groups, including teachers, researchers and policy-makers. The advantages and disadvantages of systematic reviews when compared with other forms of review, and of reviews employing the EPPI methodology have been rehearsed elsewhere (Bennett *et al.*, 2005).

Formatted: Font: Italic

The systemic review process, as developed by the EPPI-Centre, involves several stages:

- identification of review topic area and review research question or questions;
- development of *inclusion and exclusion criteria* for studies in the review (relating to, for example, aspects such as the age of students, the nature of the research design, and the reported outcomes);
- undertaking of *systematic searches* of electronic data bases and other sources for potentially relevant research studies;
- refining the search through *screening* the potentially relevant studies against the inclusion criteria;
- coding or *keywording* studies against pre-specified and agreed characteristics (some of which are generic to all EPPI reviews, whilst others are developed specifically for each review);
- production of an overview or *systematic map* of studies in the review area, that groups the studies according to their chief characteristics;

- undertaking an *in-depth review* of studies to look in detail at their design and findings and to evaluate the quality of the evidence reported.

This information is then used to make judgements about the quality of the weight of evidence presented in the study in relation to the review research question. Each of these judgements involves a decision about whether the weight of evidence in a study is *high*, *medium* or *low* Full details of the EPPI methodology may be found in the EPPI Review Group Manual (EPPI-Centre, 2002).

The EPPI review methods tend to be more tailored to quantitative research. As much of the work on small-group discussion draws extensively on qualitative approaches, the work reported here extended the EPPI review methods to draw on the guidance and framework for assessing research evidence in qualitative research studies (Spencer *et al.*, 2003). This was seen as particularly important as the majority of the studies included in the review made use of qualitative methods. As such, they ran the risk of being characterized as low in quality had judgments been informed only by the EPPI criteria.

The systematic review of studies on small-group discussions in science lessons

The main review research question was *How are small-group discussions used in science teaching with students aged 11-18, and what are their effects on students' understanding in science?* Within this, two reviews were conducted. The purpose of the first review was to gain an overview of the nature of small-group discussions used in science lessons in order to assess the levels of use, establish any patterns in use, such as in topic focus, group structure and the dynamics of group interactions. As much of the support for the use of small-group discussions is linked to their potential benefits associated with to improved understanding of science ideas, the second review focused specifically on the effects of small-group discussions on understanding.

Deleted: in relation

Deleted: students'

A note on terminology

Formatted: Font: Not Bold, Italic

The reports on small-group discussions scrutinised for the systematic reviews made it clear that the term was used in a wide variety of ways. For the purposes of the reviews, a small-group discussion was taken to be an activity that:

- involves groups of two to six students;
- has a specific stimulus (e.g. a newspaper article, video clip, prepared curriculum materials; structured teacher input);
- involves a substantive discussion task of at least two minutes (i.e. did not simply involve a student talking to a neighbour briefly about an idea or to agree and answer to a question);
- is either *synchronous* (i.e. face-to-face) or *asynchronous* (i.e. mainly IT-mediated);
- has a specific purpose (e.g. individual sense-making, or leading to an oral presentation, or to a written product).

Formatted: Bullets and Numbering

The term *understanding* has been taken to encompass understanding of science concepts, understanding of ideas about the nature of science and understanding of the methods of science.

Studies in the review

Formatted: Font: Italic

Deleted: ¶

The reviews focused on research on teaching at high school level, undertaken in the period 1980-2005, and published in English. Student age was restricted to 11-18 because this is the age range covered by the majority of reported studies, pointing to this being the school age range where the use of small-group discussions has been promoted most actively. The start date for the period of publication was selected because this was the time when the use of small-group discussions started to become more prominent in science teaching. The inclusion criteria for studies in the review were: (i) they were about the use of small-group discussions in science lessons, (ii) they involved groups of two to six students, (iii) they focused on a substantive, structured discussion task of two minutes' duration or more, and, for the second review, (iv) they addressed aspects of students' understanding in science.

Deleted: is

Deleted: ,

Deleted: and continues to be,

Deleted: , and hence is the age range covered by the majority of reported studies.

Deleted: or attitudes to science

94 studies were identified that met the above inclusion criteria. The identification of the studies is a two-step process. Firstly, systematic searches are undertaken, principally through the use of electronic search strings. Electronic searching inevitably means that large numbers of studies emerge in the initial stage, and some 2,290 studies matched the search terms. The second step involves refining the search through careful screening of abstracts (or full copies

Deleted: re

of reports if there is insufficient information in the abstract) against the inclusion criteria. The 94 studies identified were then coded for particular characteristics (*keyworded*) to produce an overview (*the systematic map*) of studies of small-group discussion interventions. In producing the map, the following characteristics of studies scrutinised included the country of study, the age/level of the students, the type of study, the science discipline of the study, constitution of discussion groups, the duration of discussion tasks, the stimulus provided, organisational features of discussion tasks, the product of discussion tasks, the research strategy used to gather data, the nature of data collected, and the outcomes reported.

An overview (systematic map) of studies on small-group discussions

This section presents a brief overview of the key features to emerge from the 94 research reports on the use and effects of small-group discussions.

Although the period of the review covered 1980-2005, over 90% were from post-1990, indicating that that the research activity area has been minimal up to fifteen years ago and has been most prolific from the late 1990s onwards. The majority of the reported work has been undertaken in North America (USA = 39%, Canada = 12%) and the UK (13%). Other work has been undertaken in Australia, The Netherlands and Germany. However, these figures need to be set in the context of the review being limited to reports published in English, though the review does include reports of studies of small-group discussions held in Bahasa Malay, Cantonese, Dutch, Finnish, French, German, Greek, Hebrew, Mandarin, Portuguese and Spanish.

Relatively little detail was given about how groups were constituted, but where this was provided, mixed-ability and friendship groups predominated. In one-third of cases, groups were deliberately constituted by the teacher, with students choosing their groups in the remaining cases. Slightly over half the studies used a group size of 3-4 students, with a further quarter of studies using pairs of students for groups. Over 80% of the studies concerned self-contained and permanent groups. The remaining studies drew on the techniques of ‘snowballing’, ‘envoying’ and ‘jigsawing’. Snowballing discussions involve progressively larger groups of students discussing a question or idea and agreeing on their views. Discussion starts with pairs, who then join together and so on. Envoying discussions involve groups of students discussing a common task. When the discussion is completed, one member of each group moves to another group to report on the discussion of the original group and to hear about the discussion in the second group. The envoy then returns to the original group to report back. Jigsawing discussions have students first working in groups where each has a different task and then moving into different groups that comprise of all the people who have focused on the same aspect of the task.

Two-thirds of the discussions took a class period or longer, with approximately two-fifths taking place in lessons described as 'science', a similar proportion in physics lessons, around one-fifth in biology lessons and only 4% in chemistry lessons. One possible explanation for this very small figure for chemistry is that most of the small-group discussions relating to exploration of difficult ideas are located mainly within physics, and those developing skills of decision-making on socio-scientific issues are more commonly placed within biology classes.

The most common stimulus material provided for the discussion tasks was printed worksheets, which were used in over two-thirds of the studies. Practical work and computer software provided the stimulus for around two-fifths and a quarter of discussions respectively. Some discussion made use of more than one stimulus. Rather surprisingly, only one study used a newspaper article as a stimulus for discussion.

In a very high proportion of the studies (94%), the main aim of the discussion task was individual understanding of the science underlying the activity, such as in a practical experiment, the preparation of a poster or a computer-based exercise, in which the learners were engaged. In the majority of cases this understanding was then shared with classmates in different ways: groups might present their findings or views orally (20%) or by way of posters (10%) or might defend their position in a whole class debate (5%). There were surprisingly few examples of written products being generated directly as an outcome of the discussion (6%).

Around three-fifths of the studies reported on evaluations of small-group discussions with the remainder providing descriptive information about the use of small-group discussions. Case studies featured prominently, with extensive use being made of video and audio recordings in order to gather detailed information about the nature of discussions. One outcome of the very labour-intensive nature of much of the data collection and analysis was that sample sizes tended to be small – very often one class or one or two groups of students within a class. Studies involving several classes, or classes in more than one school, were rare.

The chief characteristics of research on small-group discussions are summarised in Table 1.

[Table 1 about here.]

The in-depth reviews

Two in-depth reviews were conducted, covering a total of 24 of the 94 studies in the systematic map. These 24 studies were selected because (i) they reported in detail on the use of small-group discussions in science lessons (19 studies), and/or (ii) they reported evaluations of interventions aimed at developing aspects of students' understanding in science (14 studies).

Studies were rated as high (H), medium high (MH), medium (M), medium low (ML) or low (L). As quality judgements in systematic reviews are made in relation to the specific focus of each of the reviews, some studies were given different ratings in each review. Ratings were based on the extent to which the studies reported met a range of criteria relating to (i) the nature of the sample and how it was selected; (ii) the nature of any control group (for evaluations); (iii) the extent to which small-group discussions formed the main feature or variable being investigated (for evaluations); (iv) the level of detail proved about the discussion task; (v) the steps taken to establish the reliability and validity of the data collection tools and processes, and the data analysis; (vi) the trustworthiness and relatability of the data collection and analysis for qualitative data; (vii) the representativeness of the data collection situation to normal classroom situations; (viii) the quality of the reporting. No studies met all the relevant criteria, so none were judged to be high quality. The studies rated as 'medium high' met most of the relevant criteria. Studies rated at the lower end of the scale displayed one or more of the following characteristics: they provided comparatively little detail of the discussion tasks, they tended to be weaker in the reporting of steps taken to enhance the reliability and validity of data collection and analysis, and/or they were overly-descriptive at the expense of analysis and discussion. The categorisation process is described in detail in the full technical reports (see References section).

Table 2 lists the studies and their quality rating for each review.

[Table 2 about here.]

The remaining discussion focuses on the nineteen studies rated as ‘medium’ or better, as these provide the stronger evidence, and Table 3 summarises the key features of each of these studies.

[Table 3 about here.]

The detailed evidence from the in-depth reviews

The consolidated evidence presented draws primarily on the findings of the fourteen studies in the Review 1 and the twelve studies in Review 2 weighted as medium or better in overall quality, as these studies generated the better quality evidence. Of these studies, nine were common to both reviews, i.e. they focused on aspects of the use of small-group discussions and the development of understanding. The features discussed below are considered to be those that offer the strongest evidence as they emerged from three or more studies. To avoid undue repetition, evidence from both reviews is considered together under the following five headings: focus of the discussion topic, group structures and interactions, negotiating and agreeing meaning through discussion, effects on understanding, factors promoting effective discussion to enhance understanding

Focus of the discussion topic

The studies were based on a range of science topics, as Table 3 demonstrates. Seven studies addressed aspects of understanding of science topics. These focused on light (Roth and Roychoudhury, 1992), kinetic theory (Palincsar *et al.*, 1993), genetics (Finkel, 1996; Jiménez-Aleixandre *et al.*, 2000), two physics topics [shadows, floating and sinking] (Woodruff and Meyer, 1997), mechanics (Tao, 2001), and density (Kurth *et al.*, 2002). Four looked primarily at aspects of what could be termed scientific method: hypotheses on the diagnosis of disease (Richmond and Striley, 1996; Lajoie, 2001), designing controlled experiments (Sherman and Klein, 1995), and building theories and models from primary evidence on elements and bonding (Keys, 1997). Three had a specific focus on socio-scientific issues in relation to the greenhouse effect (Gayford, 1995), genetic engineering (Zohar and Nemet, 2002), and environmental science (Jiménez-Aleixandre and Pereiro-Muñoz (2002). Three involved making predictions based on evidence presented in the topics of sound (De Vries *et al.*, 2002), a range of biology topics (Lavoie, 1999), and mechanics (Tolmie and Howe, 1993). The studies by Hogan (1999a and 1999b) had no specific topic focus but were based on a series of discussion task developed by the researchers.

The studies in the in-depth review reflect the patterns noted in the systematic map, where the bulk of the discussion topics lay in the areas of physics and biology. Although there was diversity in topic focus, the common link between the discussions focusing on the development of understanding was that they all required students to draw on evidence to support a particular hypothesis, theory or point of view.

Group structures and interactions

The principal evidence on group structure and interaction came from five studies: Tolmie and Howe (1993), and Richmond and Striley (1996); Keys (1997); Hogan (1999a); De Vries *et al.*, 2002.

Group leadership emerged as crucial in promoting effective discussion. Richmond and Striley (1996) and Kurth et al. (2002) established the need for a leader to adopt an inclusive style and share tasks equitably around a group, as this promoted more substantial engagement in the discussion by a number of participants, and increased the quality of the discussion. This, in turn, permitted most members to develop their understanding. Non-inclusive leadership generated much off-task talk and engagement was generally low. Hogan (1999a) found that at least one group member had to act in a way which promoted reflection in the group for understanding of science ideas to be developed.

The evidence on assigning specific roles to students was mixed. Kurth *et al.* (2002) advocated assigning particular roles to pupils in groups as a means of achieving effective discussion. However, Richmond and Striley, (1996) and (Hogan, 1999a) reported allocating roles to have benefits when tasks were well-structured but counterproductive in poorly-structured tasks, adding to students' difficulties in engaging with the task. Hogan identified eight sociocognitive roles in group reasoning processes which took place in small-group discussions, all of which were persistent over time, and independent of particular formal roles that might have been allocated to students in advance of a discussion. Four of Hogan's roles were positive (promoter of reflection, contributor of content knowledge, creative model builder, mediator of social interaction and ideas) and four were negative (promoter of distraction, of acrimony, of simple task completion, reticent participant).

Aspects of group composition were explored in four studies (Tolmie and Howe, 1993; Keys, 1997; Hogan, 1999a; De Vries *et al.*, 2002). Tolmie and Howe found improved understanding to be independent of group composition (male, female, mixed), with biggest improvements being noted when groups contained members with a high degree of dissimilarity in their initial predictions and explanations. In common with Keys and De Vries *et al.,* Tolmie and Howe identified clear differences in interactional styles with all-male groups confronting differences in their individual predictions and explanations, whilst all-female groups searched for common features of their predictions and explanations in order to avoid conflict. Mixed groups interacted in a more constrained way than single-gender groups, though they also tended to avoid conflict and look for common patterns in contributions. Tolmie and Howe suggested that both male pairs and female groups demonstrated qualities one would want to see in the development of arguments but did not see this as a reason for promoting the use of mixed gender groups, as their study suggested the best of the all-male and all-female group interactions was lost in mixed pairs. Hogan also found that friendship groups, which were generally single-sex, functioned more effectively and promoted better development of understanding than mixed or teacher-constituted groups.

The studies suggest that both the behavioural characteristics and gender composition of groups need careful consideration if groups are to function purposefully during discussion tasks.

Negotiating and agreeing meaning through discussion

Three of the studies (Roth and Roychoudhury, 1992; Keys, 1997; Jiménez-Aleixandre *et al.*, 2000) reported in detail on the ways in which meanings were negotiated and agreed by groups.

Keys (1997) identified three common characteristics of reasoning in discussions: recognising that prior ideas (models) may be incorrect; evaluating new observations for consistency with current ideas and using evidence to modify ideas; and coordinating all mutually consistent knowledge propositions into a coherent model. A similar pattern was described by Roth and Roychoudhury (1992) who found discussions usually involved positions being stated and contested, with views either accepted or temporarily or permanently rejected as positions finally stabilised into shared meaning. Less positively, they found that students tended not to engage very often in processes which fostered meaning. Rather they would reach agreement on the basis of finding something agreeable to all group members. Agreements were often reached by one or more group members exerting authority, and on the basis of 'majority rule', rather than agreed shared understanding. These findings were echoed by Jiménez-Aleixandre *et al.* (2000) who reported that a large proportion of student talk related to what they termed 'doing the lesson', or interactions referring to the rules of the task, rather than talk related to the focus of the task. Jiménez-Aleixandre *et al.* also noted that arguments were frequently developed by a subset within the group and, though agreement was generally reached, this was often for social reasons, rather than because of agreement over the outcome of the discussion.

Whilst the above findings suggest that groups can work together to develop common understandings, they also suggest that a number of factors can influence the way in which these understandings are reached, and relative lack of engagement with the science content, coupled with the influence of students prepared to express views very strongly means that the gains in understanding may not be that high.

Effects on understanding

All the studies that looked at development of understanding reported benefits from small-group discussion work, though this finding does need to be interpreted with some caution as the majority of those undertaking the research were advocates of the approach. The following provide examples of the effects that were reported.

Roth and Roychoudhury (1992) established that group discussion over the construction of a concept map provided a vehicle for negotiation of meaning and understanding of concepts and their relationships, thus providing a structure through which students were able to learn the language patterns of science and use these to construct scientific knowledge. Tolmie and Howe (1993) reported significant improvements in students' predictions of the trajectories of falling objects through the use of a computer-based simulation. Richmond and Striley (1996) noted increasing levels of sophistication and increased use of subject knowledge in the arguments students developed in discussion of socio-scientific issues. Similarly, Zohar and Nemet (2002) reported substantial changes in the quality of student arguments in the context heredity and genetics.

Factors promoting effective discussion to enhance understanding

The factors that emerged that contributed to effective discussions and enhancement in understanding are of particular interest. Findings from four studies pointed to improvements in understanding being greatest for discussion tasks where there was dissimilarity or conflict in understanding or views. This might take the form of either *internal conflict*, or differences held by individual group members (Tolmie and Howe, 1993; De Vries *et al.*, 2002), or *external conflict* where an external stimulus presents a group with conflicting views (Tolmie and Howe, 1993; Gayford, 1995; Finkel, 1996). In some of the studies the discussion topic was selected to provide opportunities for both internal and external conflict. For example, Tolmie and Howe (1993) required students to make individual predictions about aspects of forces and motion, then engage in a task which required a joint prediction (internal conflict) and finally to compare this with an actual situation to reach an explanation of any discrepancies (external conflict). Whilst other studies did not comment specifically on the need for dissimilarity, it was clear from some of the accounts (e.g. Zohar and Nemet, 2002) that internal and external conflict were built into the discussion tasks. It seems likely that the dissimilarity in views provides a very clear and immediate focus to engage students in discussion.

Two studies offered comments on the nature of the data provided to students for the discussion. Jiménez-Aleixandre *et al.* (2000) indicated that hypothetical, unquestionable data (provided by the teacher) generates different patterns of discussion than empirical, uncertain data, perhaps gathered by students themselves, with the former leading to greater gains in understanding. In a similar vein, Roth and Roychoudhury (1992) found discussion to be more productive if students were provided with a fixed set of concepts to delimit the content of the discourse.

Three studies pointed to improved understanding when students were given specific instructions on how to construct arguments or cues to guide them in the points they needed to include (Sherman and Klein, 1995; De Vries *et al.*, 2002; Zohar and Nemet, 2002). This finding was reflected in the more general observations in two further studies that scaffolding routines, or structuring discussion through the provision of interim targets, also improved students' understanding (Palincsar *et al.*, 1993; Finkel, 1996).

Although gains in understanding were reported, the studies also suggested that students often struggled to formulate and express coherent views during small-group discussions, and demonstrated a low level of engagement with tasks. It is therefore not surprising that a number of the studies made recommendations relating to the need for students and/or teachers to be given explicit teaching in the skills associated with the development of arguments and the characteristics associated with effective group discussions.

Three studies (Richmond and Striley, 1996; Hogan, 1999a; Zohar and Nemet, 2002) recommended training for both students and teachers in the skills needed for handling and participating in group discussions. Richmond and Striley indicated that productive learning was unlikely to take place on a large scale through the use of small-group discussions until students acquire the skills associated with inclusive leadership and are thus able to foster a climate of equitable participation. Hogan (1999a) argued that guiding students towards taking constructive roles in discussions could be achieved through metacognitive training, i.e. knowledge about the nature of collaborative learning, effective group learning strategies, and awareness of what constitutes progress. Two studies (Jiménez-Aleixandre *et al.*, 2000 and Roth and Roychoudhury, 1992) recommended coaching in argumentation skills for both teachers and students.

Jiménez-Aleixandre *et al.* (2000) suggested that effective discussions are only likely to take place when linked to specific, inquiry-focused tasks where help is given to students to develop their understanding through the construction of arguments. Similarly, Roth and Roychoudhury (1992) reported that students frequently struggled with language, often making short utterances, and appeared to find it difficult to clarify their understanding through explanations, justifications and elaborations. This led them to conclude that a major outcome of their study was the recognition that understanding was only likely to be improved is students were given help in constructing arguments.

One study (Zohar and Nemet, 2002) did involve incorporating explicit instruction about argumentation into their intervention. One introductory lesson involved arguments being defined and their structure explained, together with providing examples of characteristics of good arguments. Students then practised the principles through several concrete examples. Zohar and Nemet concluded that argumentation skills were enhanced by explicit instruction and several opportunities for students to take part in discussions to help develop their skills.

Deleted: g

The nature of the evidence presented in a substantial number of the studies points to the importance of training for teachers and students. This is all the more important as many of the researchers involved in the studies were committed to the use of small-group discussions and had developed some proficiency in designing discussion tasks. This review finding also resonates with the recommendations of Levinson and Turner (2001) and Osborne *et al*. (2002) in their evaluations of two programmes focusing on teaching of socio-scientific issues, and noted earlier in this paper.

Methodological considerations

Although the primary focus of the reviews was to gather substantive data on the use of small-group discussions, some of the methodological aspects have a bearing on the nature of evidence yielded.

Positive features of the data collection included the use of multiple data sources with all studies drawing on at least two different kinds of data to increase trustworthiness. Whilst virtually all studies used video recording and/or audio recording to make verbatim records of discussions, these were supported by direct observation to record field notes, interviews, products of student tasks, such as concept maps, student questionnaires and measures of student knowledge were obtained. Although the methods used were rarely justified, the picture gained was one of studies collecting extensive data in an attempt to get as detailed a picture as possible of students' dialogue.

There were a number of limitations to the data collection. All the studies used a convenience sample for the identification of schools, in many cases using schools where access had been secured through previous involvement of the researcher. With one exception (Zohar and Nemet, 2002), the studies were based in one school and often within one class. A characteristic of much of the work was the use of retrospective sampling, i.e. data were gathered on a number of groups, but reports presented on only a sample of the groups within this, depending on characteristics of the discussion which emerged in the analysis. Such sampling methods are probably realistic for research studies fitting in with practice, and requiring extensive periods of data collection and thus a high degree of co-operation with the class teachers involved. However, retrospective sampling does confer the option on the researcher of exercising a high degree of selectivity in relation to the data presented.

Three sizeable studies (Gayford, 1995; Lavoie, 1999; Zohar and Nemet, 2002) utilised an experimental design, making comparisons between a control group who experienced a more conventional teaching approach with a group which received some form of intervention related to small-group discussion work. One study (Tolmie and Howe, 1993) specifically set up groups where gender was a variable to be explored. However, the emphasis of the majority of the studies was on describing and interpreting the nature of student discussions and their effects on students' understanding, sometimes making detailed comparisons between groups participating in their studies. Two factors may contribute to the absence of a control group in the studies. Firstly, those undertaking the research might see no need to design their studies to include a control group in what were largely qualitative and interpretative studies. Secondly, the practicalities associated with collecting and analysing extensive in-depth data from a much larger sample in order to make such comparisons would place prohibitive resource constraints on the studies.

Data analysis was characterised positively by the presentation and discussion of rich and detailed data in the form of extracts from students' discourse, with all studies adopting procedures to increase the trustworthiness of the analysis by having two or more people involved. However, given that the studies were largely gathering qualitative data, there was a surprising lack of contextual detail. Data also tended to be presented in a rather convergent manner, with few examples of data being presented which might disprove assertions or report on unintended outcomes.

Of particular interest were the two very contrasting overall analysis strategies apparent in the studies. The first strategy, adopted by the majority of the studies, was to develop grounded theory from the data through the development of categories then used to characterise the interactions between participants. Hogan (1999a) referred to her analysis as 'ethnographic interaction analysis', whilst Roth and Roychoudhury (1992) used what they described as the techniques used by anthropologists when analysing interactive behaviours. The second strategy, whose use was more limited, was to draw on existing work on discourse analysis or discourse analysis classifications. Such an approach involves trying to identify themes in what people say by looking at sentences, groups of sentences or sentence fragments. These might, for example, relate to attempts to cite others to support a view, or use of evidence to support an account of an event. Discourse analysis techniques were employed by Jiménez-Aleixandre *et al.* (2000), who drew on the work of Bloome *et al.* (1989) to do the initial coding of exchanges between students, and then used Toulmin's (1958) work on argument to classify the interactions where students were talking about science aspects in the discussion. Keys (1997) drew on elements of a framework developed by Kuhn (1993) to code students' verbal interactions relating to scientific reasoning. However, there was a notable absence of justification in the studies for the approach adopted for analysis, with the development of grounded theory appearing to be seen as an unproblematic choice in the majority of cases. It may be the case that the choice of approach reflects the personal views of the researchers on the role and purposes of data in research. However, the lack of reference made to discourse analysis techniques suggests that these approaches may be unfamiliar to some researchers working in the area of small-group discussions, which, in turn, may be limiting the nature of the analysis. There would appear to be a good case for those researching the effects of small-group discussions to gain a greater familiarity with discourse analysis techniques.

Comment [JMB1]: My understanding is that the term 'grounded theory' can be used generically to apply to a situation where researchers develop theirs about an idea from their data. Thus it does describe the process which took place in a number of the studies included in the review, even if the researcher s themselves did not employ the term.

Deleted: For example,

Comment [JMB2]: We have taken the description of the work used by the researchers in their paper.

Deleted: Both approaches have their merits.

Formatted: Font: Not Italic

Table 4 summarises the key findings to emerge from studies on the use and effects of small-group discussions in science.

[Table 4 about here.]

Conclusions and recommendations

The reviews reported here have yielded insights on both the substantive focus (small-group discussions and their effects) and on the methods employed to gather the data. The review has revealed a number of features of particular interest in relation to the use of small-group discussion work in science. It is clear from the review that a complex and interacting set of factors are involved in enabling students to engage in dialogue in a way that could help them draw on evidence to develop and articulate their understanding of science ideas.

Two particularly strong features which have emerged from the work undertaken for the review is that there is a relatively little good quality systematic research on the effects of small-group discussion work, and considerable uncertainty on the part of teachers as to what they are required to do to implement good practice. Current policy is strongly advocating the use of small-group discussion work in science, and the reviews do indicate that there could be benefits arising from this, as small-group discussion work can provide an appropriate vehicle for assisting in the development of students' understanding of science ideas. Thus teachers should be encouraged to incorporate such discussions into their teaching. However, it is also clear that small-group discussion work needs to be supported by the provision of support and guidance for teachers and students on the development of the skills necessary to make such work effective.

One feature, notable by its absence, was the dearth of systematic evidence on the effects of the use of small-group discussions on students' attitudes to their science lessons or science more widely. The absence of such data was very surprising, as the motivational effects of small-group discussions are often cited as a reason for their inclusion in science teaching.

It is clear from this review that there is considerable variation in the nature of research into small-group discussion work, particularly in relation to its focus, the clarity with which any variables being investigated are specified, the use of opportunistic samples for data collection, and the techniques used to analyse data. Particularly striking are the two very contrasting approaches to data analysis, with some studies developing grounded theory from the data, and others drawing on existing models to structure their analysis. A substantial proportion of the work also focuses on descriptive data. This can be very helpful in the early stages of a new research area. However, with increasing interest in the *effects* of small group discussions – on student learning, understanding, and attitudes – there is a need to consider what strategies and techniques lend themselves best to the gathering and analysis of data that would help explore such effects.

Taken together, these findings suggest there are four key features that need to characterise any further developments in the use of small-group discussions in science teaching and research into their effects. Firstly, some form of professional development training for teachers is highly desirable to provide guidance on how to maximise the effectiveness of small-group discussions. Secondly, further research into the effects of small-group discussions should include a consideration of the extent to which analysis of the data might benefit from established discourse analysis techniques developed in other subject areas (e.g. Barnes and Todd, 1997; Mercer and Littleton, 2007), to establish what they might have to offer work in science. Thirdly, the area would benefit from a more detailed exploration of the effects of small-group discussions on attitudinal effects. Finally, in relation to providing evidence of the *effects* of small-group discussions, there would appear to be potential benefits associated with adopting a mixed method approach to data collection, marrying in-depth qualitative data on the nature of discussions with more quantitative data on student attributes.

Deleted: would benefit from a consideration of

Acknowledgements

The work reported in this paper was funded by the Department for Education and Skills (DfES) as part of the *Evidence, Policy and Practice Initiative* (EPPI).

Electronic searches conducted for the review

Electronic searches of the following databases were conducted:

British Education Index (BEI)
Educational Resources Information Center (ERIC)
PsycINFO
Social Science Citation Index (SSCI)

References in text of report

Barnes, D. & Todd, F. (1997) Communication and learning in small groups. London: Routledge and Kegan Paul.

Bennett, J., Lubben, F., Hogarth, S. & Campbell, B. (2005) Systematic reviews of research in science education: rigour or rigidity? *International Journal of Science Education*, 27 (4), 387-406.

Bentley, D & Watts, M. (1992) Communicating in school science. London: Falmer Press.

Black, P. & Wiliam, D. (1998) *Inside the black box: raising standards through classroom assessment.* London: King's College.

Davies, P. (2000) The relevance of systematic reviews to educational policy and practice. *Oxford Review of Education*, 26, 365-378.

Daws, N. & Singh, B. (1996) Formative assessment: to what extent is its potential to enhance pupils' science being realised? *School Science Review*, 77 (281), 93-100.

Department for Education and Employment (DfEE) (1998) *The National Literacy Strategy*. London: DfEE.

Driver, R. & Bell, B. (1985) Students' thinking and the learning of science: a constructivist view. *School Science Review*, 67 (240), 443-456.

Driver, R., Asoko, H., Leach, J., Mortimer, E. & Scott, P. (1994) Constructing scientific knowledge in the classroom. *Educational Researcher*, 23 (7), 5-12.

EPPI-Centre (2002) *EPPI-Centre Review Group Manual*. London: Social Science Research Unit, Institute of Education.

Hargreaves, D. (1996) *Teaching as a research-based profession: possibilities and prospects*. London: The Teacher Training Agency (TTA).

Hillage, L., Pearson, R., Anderson, A. & Tamkin, P. (1998) *Excellence in research on schools.* Brighton: Institute for Employment Studies.

Kuhn, D. (1993) Science as argument: implications for teaching and learning scientific thinking. *Science Education,* 77 (2), 319-337.

Kyriacou, C. (1998) Essential teaching Skills (2nd edition) Cheltenham: Stanley Thornes.

Lemke, J. (1990) Talking science. New York: Ablex Publishing Corporation.

Levinson, R. & Turner, S. (2001) *Valuable Lessons: Engaging with the Social Context of Science in Schools.* London: The Wellcome Trust.

Mercer, N. & Littleton, K. (2007) Dialogue and the Development of children's thinking: a sociocultural approach. Oxford: Routledge.

Millar, R. (2006). *Twenty First Century Science*: Insights from the design and implementation of a scientific literacy approach in school science. *International Journal of Science Education*, 28 (13), 1499-1521.

Millar, R. & Osborne, J. (eds) (1998) *Beyond 2000: Science Education for the Future*. London: King's College/The Nuffield Foundation.

Millar, R., Leach, J., Osborne, J. & Ratcliffe, M. (2006) *Improving subject teaching: lessons from research in science education.* London: Routledge.

Deleted: Brighton:Institute

Newton, P., Driver, R. & Osborne, J. (1999) The place of argumentation in the pedagogy of school science. *International Journal of Science Education,* 21 (5), 553-576.

Oakley, A. (2002) Social science and evidence-based everything: the case of education. *Educational Review*, 54(3), 21-33.

OECD (2002) *Educational research and development in England: examiners' report.* Organisation for Economic Co-operation and Development. www.oecd.org/dataoecd/17/56 (accessed 2 January 2008)

Osborne, J. & Collins, S. (2001) Pupils' views of the role and value of the science curriculum. *International Journal of Science Education*, 23 (5), 441-467.

Osborne, J., Duschl, R. & Fairbrother, R. (2002) *Breaking the Mould? Teaching Science for Public Understanding*. London: The Nuffield Foundation.

Scott, P. (1998) Teacher talk and meaning-making in science classrooms: a Vygotskyan analysis and review. Studies in Science Education, 32 45-80.

Shavelson, R. & Towne, L. (2001) (eds) *Scientific enquiry in education*. Washington DC: National Academy Press.

Sinclair, J. and Coulthard, R. (1975) *Towards an Analysis of Discourse,* OUP.

Spencer, L., Ritchie, J., Lewis, J. & Dillon, L. (2003). *Quality in Qualitative Evaluation: a framework for assessing research evidence*. London: The Strategy Unit.

Sutton, C. (1992) Words, science and learning. Buckingham: Open University Press.

Toulmin, S. (1958). *The use of argument*. Cambridge: Cambridge University Press.

Torgerson, C. (2003) *Systematic Reviews*. London: Continuum.

Torgerson, C. & Torgerson, D. (2001) The need for randomised controlled trials in educational research. *British Journal of Educational Studies*, 49 (3), 316-328.

Wellington, J., & Osborne, J. (2001). *Language and literacy in science education*. Buckingham: Open University Press.

References to the full systematic review technical reports

Author 1, Author 3, Author 2, Author 4, & Author 5. (2004) A systematic review of the nature of small-group discussions in science teaching aimed at improving students' understanding of evidence. In: *Research Evidence in Education Library*. London: EPPI-

Centre, Social Science Research Unit, Institute of Education. http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=321

Author 1, Author 3, Author 2 & Author 4. (2004) A systematic review of the use of small-group discussions in science teaching with students aged 11-18, and their effects on students' understanding in science or attitude to science. In: *Research Evidence in Education Library*. London: EPPI-Centre, Social Science Research Unit, Institute of Education. http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=326

References for the studies included in the review

De Vries E, Lund K, Baker M (2002) Computer-mediated epistemic dialogue: explanation and argumentation as vehicles for understanding scientific notions. *Journal of the Learning Sciences* 11: 63-103.

Finkel EA (1996) Making sense of genetics: students' knowledge use during problem solving in a high school genetics class. *Journal of Research in Science Teaching* 33: 345-368.

Gayford C (1995) Science education and sustainability: a case-study in discussion-based learning. *Research in Science & Technological Education* 13: 135-45.

Hogan K (1999a) Sociocognitive roles in science group discourse. *International Journal of Science Education* 21: 855-882.

Hogan K (1999b) Thinking aloud together: a test of an intervention to foster students' collaborative scientific reasoning. *Journal of Research in Science Teaching* 36: 1085-1109.

Jiménez-Aleixandre MP, Pereiro-Muñoz C (2002) Knowledge producers or knowledge consumers? Argumentation and decision making about environmental management. *International Journal of Science Education* 24: 1171-1190.

Jiménez-Aleixandre MP, Rodriguez AB, Duschl RA (2000a) 'Doing the lesson' or 'doing science': argument in high school genetics. *Science and Education* 84: 757-92.

Johnson SK, Stewart J (2002) Revising and assessing explanatory models in a high school genetics class: a comparison of unsuccessful and successful performance. *Science and Education* 86: 463-480.

Keys CW (1997) An investigation of the relationship between scientific reasoning, conceptual knowledge and model formulation in a naturalistic setting. *International Journal of Science Education* 19: 957-970.

Kurth LA, Anderson CW, Palincsar AS (2002) The case of Carla: dilemmas of helping all students to understand science. *Science Education* 86: 287-313.

Lajoie SP, Lavigne NC, Guerrera C, Munsie SD (2001) Constructing knowledge in the context of BioWorld. *Instructional Science* 29: 155-186.

Lavoie DR (1999) Effects of emphasizing hypothetico-predictive reasoning within the science learning cycle on high school student's process skills and conceptual understandings in biology. *Journal of Research in Science Teaching* 36: 1127-1147.

Meyer K, Woodruff E (1997) Consensually driven explanation in science teaching. *Science Education* 81: 173-192.

Palincsar AS, Anderson C, David YM (1993) Pursuing scientific literacy in the middle grades through collaborative problem solving. *Elementary School Journal* 93: 643-658.

Richmond G, Striley J (1996) Making meaning in classrooms: social processes in small-group discourse and scientific knowledge building. *Journal of Research in Science Teaching* 33: 839-858.

Roth W-M, Roychoudhury A (1992) The social construction of scientific concepts or the concept map as conscription device and tool for social thinking in high school science. *Science and Education* 76: 531-557.

Sherman GP, Klein JD (1995) The effects of cued interaction and ability grouping during cooperative computer-based science instruction. *Educational Technology Research and Development* 43: 5-24.

Suthers D, Weiner A (1995) Groupware for developing critical discussion skills. In: Schnase JL, Cunnius EL *Proceedings of CSCL '95: The First International Conference on Computer Support for Collaborative Learning*. New Jersey: Lawrence Erlbaum Associates Inc, pages 341-348.

Tao PK (2001) Developing understanding through confronting varying views: the case of solving qualitative physics problems. *International Journal of Science Education* 23: 1201-1218.

Tao PK (2003) Eliciting and developing junior secondary students' understanding of the nature of science through a peer collaboration instruction in science stories. *International Journal of Science Education* 25: 147-171.

Tolmie A, Howe C (1993) Gender and dialogue in secondary school physics. *Gender and Education* 5: 191-209.

Tsai C-C (1999) ‘Laboratory exercises help me memorize the scientific truths’: a study of eighth graders’ scientific epistemological views and learning in laboratory activities. *Science and Education* 83: 654-674.

Williams A (1995) Long-distance collaboration: a case study of science teaching and learning. In: Spiegel SA *Perspectives from Teachers’ Classrooms. Action Research. Science FEAT (Science for Early Adolescence Teachers).*

Woodruff E, Meyer K (1997) Explanations from intra- and inter-group discourse: students building knowledge in the science classroom. *Research in Science Education* 27: 25-39.

Zohar A, Nemet F (2002) Fostering students’ knowledge and argumentation skills through dilemmas in human genetics. *Journal of Research in Science Teaching* 39: 35-62.

Table 1

An overview of the chief characteristics of research on small-group discussions

Substantive features

- The majority of the studies report work that has taken place in the USA, the UK and Canada.
- Small-group discussions were used with all ages of students in the secondary age range, though they are most widely used with the 11-16 age range.
- The majority of work focused on small-group discussions in relation to students' understanding.
- Very little research has been done on small-group discussions in relation to the teaching of chemistry.
- Typical small-group discussions involved groups of 3-4 students, with groups based on friendship ties, and the discussions lasting for at least 30 minutes.
- Typical small-group discussions had individual sense-making as their main aim (as opposed to, for example, leading to a group presentation) and use prepared printed materials as the stimulus for discussion.
- There were very few instances of students being asked to generate written products from their discussions.

Methodological features

- The most common research strategy used to gather data on small-group discussions was that of case study.
- The most popular techniques for gathering data were video- and audio-tapes of discussions, supported by observation, interviews, questionnaires and test results.

Table 2

Overview of studies included in the review, together with their quality ratings

Note 1: Quality ratings are H = high, MH = medium high, M = medium, ML = medium low, L = low.

Note 2: Some studies have different ratings for different reviews because quality judgements about weight of evidence are made in relation to the focus of each of the reviews.

	Study	Country	Review 1	Review 2
1	De Vries *et al.*, 2002	France	M	M
2	Finkel, 1996	USA	M	MH
3	Gayford, 1995	UK	-	MH
4	Hogan, 1999a	USA	MH	-
5	Hogan, 1999b	USA	M	M
6	Jiménez-Aleixandre *et al.*, 2000	Spain	MH	-
7	Jiménez-Aleixandre and Pereiro-Muñoz, 2002	Spain	M	-
8	Johnson and Stewart, 2002	USA	ML	-
9	Keys, 1997	USA	MH	MH
10	Kurth *et al.*, 2002	USA	M	-
11	Lajoie *et al.*, 2001	Canada	M	M
12	Lavoie, 1999	USA	-	M
13	Meyer and Woodruff, 1997	Canada	ML	-
14	Palincsar *et al.*, 1993	USA	ML	M
15	Richmond and Striley, 1996	USA	MH	-
16	Roth and Roychoudhury, 1992	Canada	MH	-
17	Sherman and Klein, 1995a	USA	-	H
18	Suthers and Weiner, 1995	USA	-	ML
19	Tao, 2001	Hong Kong	ML	M
20	Tolmie and Howe, 1993	UK	MH	MH
21	Tsai, 1999	Taiwan	ML	-
22	Williams, 1995	USA	-	L
23	Woodruff and Meyer, 1997	Canada	M	-
24	Zohar and Nemet, 2002	Israel	MH	M
	Total		*19*	*14*

Table 3: Details of the nature of the studies included in the review

Note 1: Quality ratings are H = high, MH = medium high, M = medium, ML = medium low, L = low.

	Study	Review	Quality rating	Sample details	Focus of study/nature of intervention	Data gathered
1	De Vries *et al.*, 2002	1 2	M M	• one class • 14 students • age 16-17 • groups of two • asynchronous discussion via computer	Physics: sound Student discussions (via computer) logged against 13 categories associated with explanation, argumentation, problem-solving and management	• student self-report diaries • log of computer dialogue
2	Finkel, 1996	1 2	M MH	• one class • 25 students • age 16-18 • groups of 3-4	Biology: genetics Students presented with two basic genetics models and data conflicting with these models. Students required to work in groups to produce a revised model	• audiotapes of group discussions • audiotapes of plenary class presentations and discussions • computer logs • individual diaries • student work
3	Gayford, 1995	2	MH	• two classes (control and experimental) from four schools • age 16 • groups of 3-4	Environmental science Students were presented with material on environmental issues and asked to reach views, distinguishing between evidence and opinion; control group covered the same material, but through teacher exposition	• pre- and post-tests of six topic questions. • self-completion questionnaire to measure motivation
4	Hogan, 1999a	1	MH	• one class • 24 students • age 13-14 • groups of 3 • mixed ability, friendship ties	No specific topic focus, but based on series of discussion tasks developed by the researcher	• one to one interview • audio and video tapes of discussions • field notes of class observations
5	Hogan, 1999b	1 2	M M	• 2 schools • 8 classes (four control and four experimental) • 163 students • age 11-16 • groups of 3-4 • heterogeneous for gender and ability	No specific topic focus, but based on series of discussion tasks developed by the researcher aimed at developing conceptual understanding and meta-cognitive skills relating to small-group discussions	• one to one interview • audio and video tapes of discussions • field notes of class observations • tests of conceptual understanding • psychological profiles
6	Jiménez-Aleixandre *et al.*, 2000	1	MH	• 1 class • 24 students • age 14-15 • groups of 4	Biology: genetics There was no specific intervention, but the class selected for the research was felt to be used to learner-centred activities and small-group discussion work	• observation • audiotapes of group discussions
7	Jiménez-Aleixandre	1	M	• 1 school	Environmental science	• observation

	and Pereiro-Muñoz, 2002			• 38 students • age 15-16 (plus some mature students) • groups of 4-6	Exploring students' ability to construct arguments and engage in decision-making about environmental processes	• audio and videotapes of group discussions • field notes • student-generated material
8	Keys, 1997	1 2	MH MH	• 1 school • 6 students • age 14-15 • groups of 3	Chemistry: elements and bonding Discussions focused on the development of reasoning strategies and discourse through a collaborative writing tasks	• videotapes of discussions • interviews with students • tests of conceptual understanding • student work
9	Kurth *et al.*, 2002	1	M	• one school • 4 students • age 11-12 • group of 4	Physics: density Material modified from the normal school module to incorporate discussion tasks	• tests of conceptual understanding • observation • self-completion questionnaire • self-completion diary
10	Lajoie *et al.*, 2001	1 2	M M	• 1 school • 2 classes • 40 students • age 14-15 • groups of 2	Biology: digestion Discussion facilitated through the use of a computer-learning environment, *Bioworld*.	• audio and video tapes of discussions • computer log of actions and decisions
11	Lavoie, 1999	2	M	• 1 school • 10 classes (5 control and 5 experimental) • 250 students • age 15-16 • groups of 3-4	Biology: several topics Topics taught in a standard way, and though a learning cycle model (exploration, term introduction, concept application)	• daily logs kept by teachers • observation • video-recordings of lessons • pre-post intervention tests of logical thinking, conceptual understanding and attitude • post intervention questionnaires to students and teachers
12	Palincsar *et al.*, 1993	1 2	ML M	• 2 schools • 9 classes • 230 students • age 11-12 • groups of 4	Chemistry: kinetic theory Discussion tasks aimed at modelling the working of scientific communities	• tests of conceptual understanding • interviews • video-recordings of selected groups • student logs
13	Richmond and Striley, 1996	1	MH	• 1 school • 1 class • 24 students • age 15-16 • groups of 4 • mixed ability and gender	Science: cholera epidemics and cystic fibrosis Aimed to explore difficulties students encounter when developing scientific arguments and how student interactions shaped arguments	• observation • self-completion student diaries • school/college records
14	Roth and Roychoudhury, 1992	1	MH	• 1 school • 6 classes	Physics: light Aimed to explore the development of student understanding	• video-recording of discussions • self-completion questionnaire

				• 148 students (but only one group studied in detail) • age 15-17 • groups of 3-4	through engaging in the process of developing concept maps	• concept maps generated in discussions
15	Sherman and Klein, 1995	2	H	• 1 school • number of classes unspecified • 231 students • age 13-14 • groups of 2	Science: investigations Computer programme about designing controlled experiments	• observation • self-completion questionnaire • pre- and post-tests of understanding
16	Tao, 2001	1 2	ML	• 1 school • 1 class • 16 students • age 17-18 • groups of 2	Physics: several topics Multiple solutions presented to students to see if discussions improved their understanding	• pre- and post-tests of understanding • individual interviews • audio tapes of discussions
17	Tolmie and Howe, 1993	1 2	MH MH	• 1 school • number of classes unspecified • 82 students • age 12 to 15 • groups of 2	Physics: forces and motion Exploration of gender differences in group discussions	• Assessment of understanding • Video-tapes of discussions • Psychological test • Computer record of joint predictions
18	Woodruff and Meyer, 1997	1	M	• 1 school • 3 classes • sample size not stated – probably one class • age 11-13 • groups of 3-4	Physics: shadows, and floating and sinking. Findings reported from three studies on interactions in discussions within and between groups	• audio-recordings dins of discussions • field notes
19	Zohar and Nemet, 2002	1 2	MH M	• 2 school • 9 classes (5 control, 2 experimental) • 186 students • age 13-14 • groups of 5-7	Biology: genetics Exploration of the effects of a unit teaching argumentation skills	• multiple-choice test of understanding • pre- and post-test of argumentation skills • student worksheets • audio-tapes of four discussions

Table 4

Summary of key findings on the use and effects of small-group discussions

Substantive features

- There is considerable diversity in the topics used to promote small-group discussions.
- Students often struggle to formulate and express coherent arguments.
- Students often demonstrate a low level of engagement with tasks.
- Groups function more purposefully when specifically constituted such that differing views are represented, and improvements in understanding are greatest where there is initial *dissimilarity* in understanding of the science ideas associated with the discussion task.
- Groups function more purposefully, and students' understanding improves, when the stimulus used to promote discussion involves both internal and external conflict, i.e. where a diversity of views and/or understanding are represented within a group (internal conflict) and where an external stimulus presents a group with conflicting views (external conflict).
- Groups function more purposefully, and students' understanding improves, when some form of training is provided for students on aspects of small-group discussion work, and when help in structuring discussions is provided in the form of 'cues'.
- Single sex groups function more purposefully than mixed sex groups, though improvements in understanding are independent of gender composition of groups.
- Group leaders able to adopt an inclusive style, and one which promoted reflection, are the most successful in achieving engagement with the task.
- Incorrect or inadequate prior knowledge hinders development of students' understanding through small-group discussion.
- Teachers and students need to be given explicit teaching in the skills associated with the development of arguments and the characteristics associated with effective group discussions. The effectiveness of small-group discussions, and their effects on students' understanding of evidence, is linked more strongly to the provision of such guidance, rather than to any particular type of stimulus material.
- Little systematic data has been gathered on the effects of small-group discussions on students' affective responses to science.

Methodological features

- All studies generated large data sets, and used multiple data sources to enhance the reliability and/or trustworthiness of data.
- With one exception all the studies were based on single schools, many on single classes, with schools being identified through convenience sampling.
- Relatively few studies made use of experimental designs.
- Two very contrasting approaches to analysis were adopted: one developing theory from the data ('grounded theory') and the other using established techniques for discourse analysis.